Documentos

ISSN 1982-5390
Outubro, 2008
74

Manual de utilização do programa INTERGEN - Versão 1.0[1] em estudos de genética quantitativa animal

Autor

Fernando Flores Cardoso
Pesquisador A da Embrapa Pecuária Sul

Sumário

Manual de utilização do programa INTERGEN - Versão 1.0[1] em estudos de genética quantitativa animal

Fernando Flores Cardoso

Introdução

O programa computacional INTERGEN foi desenvolvido pela Embrapa Pecuária Sul em linguagem Fortran 90/95 com capacidade de implementar modelos hierárquicos de Bayes e estimar seus parâmetros por meio de métodos Monte Carlo via Cadeias de Markov (MCMC) (SORENSEN; GIANOLA, 2002). Esses modelos contemplam uma diversidade de situações abordadas em estudos de genética quantitativa, utilizando dados de desempenho de animais domésticos, incluindo: incerteza de paternidade na presença de acasalamentos com reprodutores múltiplos, estrutura populacional com múltiplas composições raciais – populações cruzadas, interação genótipo-ambiente, heterogeneidade de variância residual e robustez a dados extremos.

O programa utiliza bibliotecas para geração de números aleatórios e para operações com matrizes esparsas e foi desenvolvido a partir de programas de implementação da metodologia dos modelos mistos (HENDERSON, 1984)e MCMC desenvolvidas por Misztal et al. (2002), disponíveis no site (http://nce.ads.uga.edu/~ignacy/).

[1]Financiado pela Embrapa (Projeto 03.04.3.26), CNPq (Processo 481565/2004-3) e Fapergs (Processo 04/0808.4)

O objetivo deste manual é descrever o programa INTERGEN e prover seus usuários com um conjunto de instruções, dicas e exemplos, que permitam a utilização do programa INTERGEN para análises genéticas quantitativas de dados de animais domésticos. Este manual está estruturado em quatros seções: Seção I descreve a utilização do programa e os arquivos necessários; Seção II é dedicada à preparação, formatação e consistência dos dados de desempenho animal e pedigree de acordo com os inputs necessários ao programa INTERGEN; Seção III apresenta dicas para rodar o programa, e finalmente na Seção IV são descritos os arquivos de saída do programa (outputs) e detalhes para análises pós-MCMC e interpretação dos resultados.

Seção I - Descrição do Programa

O programa INTERGEN, apesar de permitir modelagem complexa, possui uma interface bastante simples, sendo controlado por um arquivo de parâmetros (“parameterfile”), o qual contém as informações sobre os arquivos de dados e de pedigree, sobre os efeitos no modelo e sobre a cadeia MCMC a ser implementada. As descrições apresentadas a seguir são baseadas em uma versão compilada para Windows, acionada através de uma janela do Prompt de Comando do DOS. Entretanto, o programa pode ser compilado e rodado também em ambiente Linux e Unix, tanto em estações de trabalho, quanto computadores pessoais.

Uma vez acionado, o programa INTERGEN solicita ao usuário a especificação de um arquivo de parâmetros, que deve ser preparado previamente em texto não formatado (arquivo em ASCII, como, por exemplo, os arquivos com extensão .txt), sendo esse arquivo de parâmetros a interface que controla o programa.

1.1. Arquivo de parâmetros

O arquivo de parâmetros apresenta um formato geral, descrito com detalhes abaixo, com seções definidas por cabeçalhos em letras maiúsculas em negrito, que devem aparecer exatamente como descritos abaixo e numa única linha no arquivo ASCII – embora aqui possam aparecer em duas linhas por limitação de tamanho de linha no presente documento, e campos em itálico que são alterados pelo usuário para controlar o pro-

grama. Na seção abaixo comentários são feitos entre parênteses.

Título de descrição da análise em formato livre
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
(cabeçalho de seção obrigatória)
número total de ciclos MCMC, período de aquecimento, intervalo para salvar amostras (thinning)
SEED (cabeçalho de seção obrigatória)
número inteiro (semente para iniciar o processo de geração de variáveis aleatórias, o que permite repetir exatamente a mesma cadeia)
RESTART: Y/N? [CYCLE_TO_RESTART] (cabeçalho de seção obrigatória)
Informa se análise é continuidade de cadeia já em andamento? (**y** = continuar cadeia, **n** = iniciar nova cadeia), *ciclo para continuar a cadeia* (caso seja recomeço)
DATAFILE NAME N_RECORDS (cabeçalho de seção obrigatória)
nome do arquivo de dados, número de registros neste arquivo
NUMBER_OF_TRAITS (cabeçalho de seção obrigatória)
número de características respostas na análise
NUMBER_OF_EFFECTS (cabeçalho de seção obrigatória)
número de efeitos no modelo
OBSERVATION(S) (cabeçalho de seção obrigatória)
coluna(s) no arquivo de dados onde estão os valores da(s) variável(is) resposta
(uma coluna por variável)
WEIGHT(S) (cabeçalho de seção obrigatória)
coluna(s) onde estão as ponderações para a(s) variável(is) resposta
(uma coluna por variável; em branco para sem ponderação)
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED] (cabeçalho de seção obrigatória)
(Seção dos efeitos no modelo com uma linha para cada efeito)
coluna no arquivo de dados onde estão os valores do efeito (uma coluna por variável resposta), *número de níveis, tipo de efeito (***cross** = *classificatório,* **cov** = *covariável,* **unknowncov** = *efeito classificatório a ser usado como covariável para normas de reação via regressão aleatória,* **rnorm** = *normas de reação via regressão aleatória,* **ram** = *efeito de modelo animal reduzido com reprodutores múltiplos), salvar amostras? (***y** = *sim,* **n** = *não) posição na qual o efeito está aninhado* (opcional; uma coluna por variável resposta)

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF (cabeçalho de seção obrigatória)
tipo de pressuposição sobre a distribuição dos resíduos (uma de cinco opções: **homogeneous**, **structural**, **student_t**, **struct_student_t**, **slash** or **struct_slash**), *grau de confiança nos valores a priori*
METROPOLIS_STEP_OF_STRUCTURAL_EFFECTS: ROUNDS_WITHIN_CYCLE TUNING_SKIP (cabeçalho de seção opcional - somente se houver efeito estrutural nos resíduos)
número de passos do algoritmo de Metropolis-Hastings (MH) dentro de cada ciclo, intervalo de ciclos para ajustar a variância da distribuição de propostas no algoritmo de MH durante o período de aquecimento (opções avançadas do modelo estrutural).
NUMBER_OF_STRUCTURAL_EFFECTS (cabeçalho de seção opcional - somente incluir se tipo de resíduo for estrutural)
número de efeitos para a variância residual
STRUCTURAL_EFFECTS: LINE_FROM_EFFECTS_SECTION SAVE_SAMPLES? (cabeçalho de seção opcional - somente incluir se tipo de resíduo for estrutural - efeitos do modelo para variância residual com uma linha para cada efeito)
linha da seção de efeitos no modelo EFFECTS: que contém o efeito a ser modelado na variância residual (deve-se incluir o número de ordem da linha, isto é, se desejamos o efeito da primeira linha da seção EFFECTS: modele a variância residual deve usar 1) *salvar amostras? (***y** *= sim,* **n** *= não)*
RESIDUAL_PRIOR_(CO)VARIANCES (cabeçalho de seção obrigatória)
valores a priori para a matriz de covariância residual
RANDOM_GROUP (cabeçalho de seção opcional - especifica quais dos efeitos do modelo na seção EFFECTS são aleatórios e quais as pressuposições sobre distribuições desses efeitos)
linha(s) da seção de efeitos no modelo EFFECTS: (pode ser mais de uma linha em caso de efeitos aleatórios correlacionados - por ex. direto e materno; efeitos correlacionados devem ser consecutivos na seção *EFFECTS:*)
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF (cabeçalho de seção opcional – necessariamente acompanha a seção *RANDOM_GROUP*)
tipo de efeito aleatório (uma das seguintes possibilidades: **diagonal**, **add_animal**, **add_sire**, **add_an_ms**, **add_an_mb** ou **diag_mb**) *e grau de confiança no valor a priori*
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS] (cabeçalho de seção opcional – necessariamente acompanha a seção

RANDOM_GROUP)
nome do arquivo de pedigree, número de animais, número de grupos genéticos e de raças, se for análise multirracial, no pedigree (pode ficar em branco, p.ex. no caso de efeito diagonal)
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION] [DIRICHLET_PRIORS]
(Seção a ser incluída somente se o tipo de efeito aleatório = **add_an_ms** ou, **add_an_mb**, que define inferência na incerteza de paternidade quando se tem reprodutores múltiplos)
número máximo de possíveis pais para fazer inferência (usar 0 para análise sem reprodutores múltiplos – RM - e 1 para utilizar a matriz de parentesco médio, que também será usada quando o tamanho do grupo de RM exceder o número máximo de possíveis pais para fazer inferência), *nome do arquivo de RM* (opcional, somente se houverem RM) *número de linhas no arquivo de RM* e *valor para hiperparâmetros da distribuição a priori Dirichlet das probabilidades de paternidade* (se = 0 não usa a distribuição Dirichlet, se >0 , usar o valor especificado para todos os parâmetros alfa da distribuição Dirichlet em todos os grupos de RM, e se < 0, usa valores definidos pelo usuário no arquivo de RM)
METROPOLIS_STEP_OF_MULTIBREED_(CO)VARIANCES: ROUNDS_WITHIN_CYCLE TUNING_SKIP (cabeçalho de seção opcional - somente se houver efeito aleatório multirracial – RANDOM_TYPE = **add_an_mb** ou **diag_mb**)
número de passos do algoritmo de Metropolis-Hastings (MH) dentro de cada ciclo, intervalo de ciclos para ajustar a variância da distribuição de propostas no algoritmo de MH durante o período de aquecimento (opções avançadas do modelo multirracial).
(CO)VARIANCES (cabeçalho de seção opcional – necessariamente acompanha a seção *RANDOM_GROUP*)
valor a priori para a matriz de covariância para os efeitos correlacionados por resposta

A seção RANDOM-GROUP e as subseqüentes são repetidas para cada efeito ou grupo de efeitos aleatórios.

Comentários podem ser incluídos no arquivo de parâmetros antes de cada seção ou após a descrição do campo usando #.

1.2. Alguns detalhes das seções:

DATAFILE NAME N_RECORDS

Arquivo de dados: os arquivos de dados são arquivos texto não formatado (ASCII), com dados para todos os efeitos no modelo em colunas, com pelo menos um espaço em branco entre colunas, e uma linha para cada registro. No caso de modelo animal, devem conter a ID do animal ao qual o registro pertence, de acordo com a codificação no arquivo de pedigree. Todos os efeitos classificatórios devem ser recodificados de 1 ao número de classes, para minimizar o uso de memória. Não existe uma ordem necessária dos efeitos no arquivo, entretanto, por questão de organização, recomenda-se o seguinte padrão: primeira coluna com ID do animal, a seguir ID da mãe se houver efeito materno, depois demais efeitos classificatórios, seguidos de covariaveis e finalmente as variáveis respostas. Dados perdidos devem ser codificados com 0.

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT SAVE_SAMPLES [EFFECT NESTED]

Nesta seção, a opção [EFFECT NESTED] é utilizada para especificar efeitos de regressão aleatória e de normas de reação, onde posição no arquivo de dados = a posição da covariável (uma coluna por variável resposta), número de níveis = número de níveis do efeito aninhado (p.ex. efeito de animal), tipo de efeito = cov para modelo de regressão aleatória e rnorm para modelo de normas de reação com covariaveis desconhecidas, salvar amostras? (y = sim, n = não) e posição na qual o efeito está aninhado (é por. ex. a posição do efeito animal no arquivo de dados (uma coluna por variável resposta). Por exemplo, com os valores do polinômio da coluna 4 dos arquivo de dados, animal na coluna 1 com 1000 níveis e não salvando amostras (que é a opção recomendada neste caso), a linha para o efeito fica assim:
4 1000 cov n 1

Modelos diferentes para cada característica podem ser especificados. Valor 0 para a posição do efeito no arquivo de dados é usado para codificar a ausência do efeito para uma característica (variável). Usando um exemplo de Misztal (2007), se a variável 1 tem grupo de contemporâ-

neos (GC) na posição 1 com 1000 níveis e variável 2 tem GC na posição 7 com 2000 níveis, isto pode ser codificado das seguintes formas no seção EFFECTS: do arquivo de parâmetros:
1 0 1000 cross n n
0 7 2000 cross n n
ou, em uma linha como:
1 7 2000 cross n n
onde o número de níveis é máximo entre os dois.

No caso de modelo animal reduzido (tipo de efeito = ram) esse(s) efeito(s) deve(m) ser o(s) primeiro(s) a aparecer(em) na lista da seção EFFECTS:

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF

Os tipos de pressuposições possíveis são seis e são descritos em detalhe por Cardoso et al. (2005). Sucintamente, tem-se:

Normal homoscedástico (homogeneous): Distribuição normal (gaussiana) com variância homogênea $e \sim N(0,S^2_e)$ – que é a pressuposição usual;

<u>**IMPORTANTE:**</u> Esse é a única opção para análises multivariadas. As seguintes são validas (e foram testadas) somente para análises univariadas.

Normal heteroscedástico (structural): Distribuição normal com variância heterogênea $e_{(i)} \sim N(0,S^2_{e(i)})$, onde $S^2_{e(i)}$ é a variância específica dos erros $e_{(i)}$ de uma combinação de subclasses *i* definida por um modelo estrutural;

Student *t* homoscedástico (student_t): Distribuição Student t com variância homogênea $e_{(i)} \sim N(0,S^2_e/w_{(i)}) \rightarrow y_{(i)} \sim t_{(v)}(0,V^2_e)$ - modelo robusto;

Student *t* heteroscedástico (struct_student_t): Distribuição Student t com variância heterogênea $e_{(i)} \sim N(0,S^2_{e(i)}/w_{(i)}) \rightarrow y_{(i)} \sim t_{(v)}(0,V^2_{e(i)})$ - modelo estrutural robusto;

Slash homoscedástico (slash): Distribuição Slash com variância homogênea $e_{(i)} \sim N(0,S^2_e/w_{(i)}) \rightarrow y_{(i)} \sim Slash_{(v)}(0,V^2_e)$ - modelo robusto;

Slash heteroscedástico (struct_slash): Distribuição Slash com variância heterogênea $e_{(i)} \sim N(0,S^2_{e(i)}/w_{(i)}) \rightarrow y_{(i)} \sim Slash_{(v)}(0,V^2 e_{(i)})$ - modelo estrutural robusto.

Para os graus de confiança nos valores a priori da(s) (co)variância(s) dos efeitos aleatórios, quanto maior o valor especificado, maior será a influencia dos valores a priori nos resultados da análise. Especificando-se o número de variáveis na análise é equivalente a uma priori difusa pouco informativa e, portanto, a inferência é basicamente a partir da informação dos dados.

RANDOM_GROUP

Efeitos aleatórios e arquivos de pedigree

Os diferentes tipos de efeitos aleatórios requerem diferentes arquivos de pedigree. De forma geral, o arquivo de pedigree deve ter uma linha por animal com pelo menos três colunas. Todos os animais devem constar no arquivo de pedigree identificados de 1 ao número total de animais no arquivo, incluindo os animais base sem pais conhecidos. Não há necessidade de os pais terem número de ID menor que os filhos, exceto para o modelo animal reduzido. Quando pai, mãe ou ambos forem desconhecidos devem ser identificados com 0. Outras colunas serão necessárias dependendo do tipo de efeito aleatório associado com o pedigree em questão, conforme descrito a seguir:

Efeito não correlacionado (diagonal): Não requer arquivo de pedigree, pois os níveis não são correlacionados dentro de variável. Ex: grupos contemporâneos aleatórios e efeito de ambiente permanente da vaca.

Efeito de aditivo de touro (add_sire):
número do animal, número do pai, número do avô materno

Efeito aditivo de animal (add_animal):
número do animal, número do pai, número da mãe

Efeito aditivo animal com reprodutores múltiplos e grupos genéticos (add_an_ms) :
*número do animal, número da mãe, número do pai (ou -1 se for animal sem pai conhecido),1/d, indicador se o animal tem progênie no arquivo (*1=sim, 0=não)

Neste caso, d = proporção da variância genética que é atribuída à segregação mendeliana para o animal em questão. A quantidade d é uma função do coeficiente de consangüinidade e do conhecimento dos pais corretos. Para animais não consangüíneos e com ambos os genitores conhecidos d = 0,5.

Os animais que são pais devem ter necessariamente numeração menor do que os que aparecem apenas com registro próprio sem descendência.

Por exemplo, para um animal com pais desconhecidos *d* = 1 -> 100% variância genética para segregação mendeliana e pai de outros indivíduos, tem-se:
23 0 0 1 1
já para um animal com pais conhecidos e não consangüíneo *d* = 0,5 (e consequentemente 1/d=2) -> 50% variância genética para segregação mendeliana e pai de outros indivíduos, tem-se:
24 5 3 2 1
e para animal não consangüíneo com um dos pais conhecido e outro desconhecido *d* = 0,75 (1/d=1,333) -> 75% variância genética para segregação mendeliana e sem progênie, tem-se:
25 9 -1 1.3333333333 0

Na presença de incerteza de paternidade para o efeito **add_an_ms** é necessário também um arquivo de reprodutores múltiplos, onde são indicados quais os possíveis pais para cada indivíduo com paternidade incerta, que é indicado por -1 na coluna do número do pai do arquivo de pedigree. Esse arquivo tem a seguinte estrutura para cada animal:
número de possíveis pais
identificação do pai 1
probabilidade a priori do pai 1
...
identificação do pai n
probabilidade a priori do pai n

Por exemplo, para um animal com 3 pais possíveis com iguais probabilidades de paternidade a priori, tem-se:
3
683
0.3333333333
668
0.3333333333
235
0.3333333333
onde 683, 668 e 235 são as ID dos possíveis pais.

O programa INTERGEN cada vez que encontra um animal com incerteza de paternidade no arquivo de pedigree (indicado por -1 na terceira coluna), busca no arquivo de RM quais os seus possíveis pais e, portanto, neste arquivo os pais candidatos de cada animal devem aparecer na mesma ordem que esses animais com paternidade incerta são relacionados no arquivo de pedigree. A estrutura descrita acima deve ser repetida para cada um desses animais. Também a identificação dos pais deve ser consistente com a do arquivo de pedigree.

Efeito aditivo animal multirracial com reprodutores múltiplos e grupos genéticos (add_an_mb)
número do animal, número do pai, número da mãe, proporção da raça 1,..., proporção da raça n.

Para o modelo multirracial o arquivo de pedigree deve conter também a composição racial do animal, especificada pela proporção de cada uma das raças. Por exemplo, para um animal número 23, com pais desconhecidos, 5/8 raça 1 e 3/8 raça 2 o seu registro será:
23 0 0 0.625 0.375

Grupos genéticos (fixos) são especificados atribuindo a eles números maiores que o número de animais no arquivo de pedigree. Por exemplo, para associar o animal acima com um grupo genético 101, sendo que número de animais = 100, seu registro ficaria:
23 101 101 0.625 0.375

Igual ao efeito add_an_ms, na presença de incerteza de paternidade, é necessário um arquivo de reprodutores múltiplos, conforme descrito no item acima.

Efeito não correlacionado multirracial (diag_mb): Não requer arquivo de pedigree.

(CO)VARIANCES

A estrutura da matriz de (co)variâncias (G) permite múltiplos efeitos aleatórios e múltiplas características. Supondo um exemplo com dois efeitos (a e m) e duas características (1 e 2), tem-se a seguinte matriz simétrica, onde aparecem blocos diagonais de cada efeito com variâncias e covariâncias entres as diferentes características e blocos fora da diagonal com as covariâncias entre os diferentes efeitos e características:

$$G = \begin{bmatrix} \sigma^2_{a_1} & \sigma_{a_1a_2} & \sigma_{a_1m_1} & \sigma_{a_1m_2} \\ \sigma_{a_1a_2} & \sigma^2_{a_2} & \sigma_{a_2m_1} & \sigma_{a_2m_2} \\ \sigma_{a_1m_1} & \sigma_{a_2m_1} & \sigma^2_{m_1} & \sigma_{m_1m_2} \\ \sigma_{a_1m_2} & \sigma_{a_2m_2} & \sigma_{m_1m_2} & \sigma^2_{m_2} \end{bmatrix}$$

Para o caso de modelo multirracial onde as variâncias são específicas por raça e existe variância da segregação entre raças (LO et al., 1993), a matriz é especificada em grandes blocos para raças na diagonal e para segregação entre raças fora dela, conforme o exemplo abaixo, no qual os mesmos efeitos e características acima, são especificados para duas raças (b e c):

$$G^* = \begin{bmatrix} \sigma^2_{a_1b} & \sigma_{a_1a_2b} & \sigma_{a_1m_1b} & \sigma_{a_1m_2b} & \sigma^2_{a_1S_{bc}} & \sigma_{a_1a_2S_{bc}} & \sigma_{a_1m_1S_{bc}} & \sigma_{a_1m_2S_{bc}} \\ \sigma_{a_1a_2b} & \sigma^2_{a_2b} & \sigma_{a_2m_1b} & \sigma_{a_2m_2b} & \sigma_{a_1a_2S_{bc}} & \sigma^2_{a_2S_{bc}} & \sigma_{a_2m_1S_{bc}} & \sigma_{a_2m_2S_{bc}} \\ \sigma_{a_1m_1b} & \sigma_{a_2m_1b} & \sigma^2_{m_1b} & \sigma_{m_1m_2b} & \sigma_{a_1m_1S_{bc}} & \sigma_{a_2m_1S_{bc}} & \sigma^2_{m_1S_{bc}} & \sigma_{m_1m_2S_{bc}} \\ \sigma_{a_1m_2b} & \sigma_{a_2m_2b} & \sigma_{m_1m_2b} & \sigma^2_{m_2b} & \sigma_{a_1m_2S_{bc}} & \sigma_{a_2m_2S_{bc}} & \sigma_{m_1m_2S_{bc}} & \sigma^2_{m_2bc} \\ \sigma^2_{a_1S_{bc}} & \sigma_{a_1a_2S_{bc}} & \sigma_{a_1m_1S_{bc}} & \sigma_{a_1m_2S_{bc}} & \sigma^2_{a_1c} & \sigma_{a_1a_2c} & \sigma_{a_1m_1c} & \sigma_{a_1m_2c} \\ \sigma_{a_1a_2S_{bc}} & \sigma^2_{a_2S_{bc}} & \sigma_{a_2m_1S_{bc}} & \sigma_{a_2m_2S_{bc}} & \sigma_{a_1a_2c} & \sigma^2_{a_2c} & \sigma_{a_2m_1c} & \sigma_{a_2m_2c} \\ \sigma_{a_1m_1S_{bc}} & \sigma_{a_2m_1S_{bc}} & \sigma^2_{m_1S_{bc}} & \sigma_{m_1m_2S_{bc}} & \sigma_{a_1m_1c} & \sigma_{a_2m_1c} & \sigma^2_{m_1c} & \sigma_{m_1m_2c} \\ \sigma_{a_1m_2S_{bc}} & \sigma_{a_2m_2S_{bc}} & \sigma_{m_1m_2S_{bc}} & \sigma^2_{m_2bc} & \sigma_{a_1m_2c} & \sigma_{a_2m_2c} & \sigma_{m_1m_2c} & \sigma^2_{m_2c} \end{bmatrix}$$

Note que a matriz G^* acima é uma maneira de especificar para o programa todos os componentes de variância das raças b e c e da segregação entre as essas raças S_{bc}, mas <u>não</u> é a matriz de covariância multirracial.

Embora as matrizes acima sejam simétricas, todos seus os elementos devem ser especificados na seção *(CO)VARIANCES.*

Seção II – Preparação, formatação e consistência dos dados

Existem diferentes formas e seqüências possíveis para preparar os dados para rodar o programa INTERGEN. A seguir é detalhado um exemplo realizado com o programa SAS (SAS INSTITUTE, 2004) *para preparar os dados para rodar um modelo animal (Anexo 1) e modelo de normas de reação (Anexo 2) para estudos de interação genótipo-ambiente*, que pode ser usado como referência para o procedimento de preparação, formatação e consistência dos dados.

Os passos adotados foram os seguintes:

2.1. Leitura dos dados até o teste de conectabilidade (Rotina em Pré-INTERGEN1.sas no Anexo 4):

2.1.1. Ler os dados originais para o programa que fará a preparação.

2.1.2. Formar grupos contemporâneos para serem usados como descritores dos ambientes nos quais os animais foram criados. De acordo com a estrutura da população, características de manejo e as informações disponíveis sobre o conjunto de dados, esses grupos devem ser formados para agrupar os animais que tiveram um ambiente comum ou equivalente para expressar seu potencial produtivo. Por exemplo: rebanho-ano-estação em gado de leite e rebanho-ano-estação-código de manejo-data da pesagem-sexo para gado de corte. Preferentemente, não devem se formar grupos muito pequenos.

2.1.3. Verificar e formar, se necessário, identificações únicas para os animais, por exemplo, incluindo número original, rebanho e ano de nascimento na identificação. É necessário que os animais tenham a mesma identificação quando aparecem como produtos e depois como pais de outros animais, para combinar as informações de desempenho individual e da progênie no modelo animal.

2.1.4. Formar outras variáveis a serem utilizadas na análise, a partir dos dados originais. Por exemplo, idade da vaca ao quadrado (IDV2), idade à desmama (IDD), a partir das datas de nascimento e desmama, etc.

2.1.5. Fazer estatísticas descritivas: Freqüência para dados discretos (variáveis classificatórias) e médias, mínimos, máximos, desvios padrão, etc, para variáveis contínuas (variáveis resposta e covariaveis). Checar inconsistências: pesos e idades muito pequenos ou muito grandes, valores de classes impossíveis, etc.

2.1.6. Consistência: eliminar registros fora dos padrões de peso, idade, etc.; em caso de modelo não robusto, como no caso de pressuposição de distribuição normal dos resíduos, eliminar registros extremos dentro do grupo de contemporâneos (p.ex. $>$ 3,5 desvios padrões abaixo ou acima da média do grupos de contemporâneos); eliminar grupos contem-

porâneos com poucos registros (p. ex. 5 ou menos).

2.1.7. Teste de conectabilidade de GC baseado no número total de laços genéticos (mínimo 10), baseado em um modelo animal, usando o programa AMC (ROSO, SCHENKEL, 2006). O programa junto com as instruções pode ser obtido diretamente dos autores. O programa permite varias opções para testar conectabilidade, onde se pode escolher o modelo (touro e vaca, animal, etc.), o número mínimo de laços genéticos e observações para considerar o grupo contemporâneo conectado (ver instruções do programa para maiores detalhes). O programa localiza o grupo com maior número de conexões genéticas e depois todos os demais grupos a ele conectados (arquipélago 1). Depois procura o grupo com maior número de conexões entre os não conectados ao arquipélago 1 e os demais conectados a ele para formar o arquipélago 2 e assim por diante até que permanecem somente grupos completamente desconectados. A rotina prepara o arquivo de entrada para o programa AMC, de acordo com o formato necessário em linguagem Fortran:
FORMAT(3a10,21x,a7,a1,24x,a1).

Para rodar o programa AMC: O primeiro programa (Pre_INTERGEN1.sas) gera um arquivo chamando infile.txt. Esse arquivo deve ser colocado em um diretório junto com o programa AMC (amcw1.exe). Abrir uma janela do prompt de comando do DOS (no menu iniciar -> programas -> acessórios) e ir até o diretório onde está o programa. Digitar: amcw1 > saida_amcw1 <enter>. O AMC irá rodar, gerar seus arquivos de resultados e toda a saída de tela será armazenada no arquivo saida_amcw1.

2.2. Leitura da saída do programa AMC e preparação de arquivos para rodar o modelo animal (Rotina Pré-INTERGEN2.sas - Anexo 5):

2.2.1. Lê o arquivo de saída do programa AMC para identificar os animais conectados. Para os propósitos desta análise, foi usado um modelo animal (todos os laços genéticos através da matriz de parentesco foram considerados na conectabilidade) e um mínimo de 10 laços foram necessários para considerar o grupo conectado e somente os grupos de arquipélago 1 foram mantidos para a análise.

2.2.2. Gera o arquivo de pedigree. Produtos, touros e vacas são recodificados de 1 ao número de animais, sem a preocupação de que pais tenham números de ID menores que os filhos, pois isto não é necessário para o modelo animal no INTERGEN. Pais desconhecidos, bem como reprodutores múltiplos são codificados com 0 (não foi usado modelo de incerteza de paternidade). Um código de acordo com o número de pais desconhecidos é gerado, 1 para nenhum (pai e mãe conhecidos), 2 para 1 pai (ou mãe) desconhecido e 3 para os dois pais desconhecidos (animais base). No exemplo adotado, o pedigree é aumentado através de inclusão dos pais dos touros que aparecem no sumário de touros da ANC, para aumentar os laços genéticos.

2.2.3. Gera arquivo de dados para o modelo animal. Renumera os efeitos classificatórios (1 a n), no exemplo somente grupos de contemporâneos são usados como classificatórios, além do animal que já está recodificado. As variáveis gravadas no arquivo devem ser escolhidas de acordo com o modelo a ser adotado, incluindo o animal (e mãe, se houver efeito materno), efeitos fixos e aleatórios e variáveis respostas. Opcionalmente, uma coluna somente com 1's pode ser usada para ajustar uma constante no modelo.

2.2.4. Estatísticas descritivas do arquivo de dados são calculadas para verificar se os dados foram gerados corretamente e para serem usadas na descrição dos dados.

2.2.5. Por fim, neste arquivo, o PROC MIXED dos SAS (SAS INSTITUTE, 2004) é rodado para verificar a significância e pertinência do modelo misto proposto, antes de rodar o programa INTERGEN.

A seguir rodar o programa INTERGEN com um modelo animal (ver detalhes Seção III e Anexo 1), para obter as estimativas dos efeitos médios de ambiente (no exemplo, soluções para grupos de contemporâneos) e também para servir de base de comparação para a análise de normas de reação.

2.3. Leitura das soluções do modelo animal e geração do arquivo para rodar modelo de normas de reação (Rotina Pré-INTERGEN3.sas - Anexo 6):

2.3.1. Lê arquivo de dados gravado pela rotina Pre_INTERGEN2.sas.

2.3.2. Lê arquivo "solutions" com as soluções do programa INTERGEN para os efeitos fixos e aleatórios do modelo animal. É preciso especificar qual entre os efeitos contém as soluções para os efeitos ambientais - neste caso são os grupos de contemporâneos - efeito 3 - na terceira linha da secção EFFECTS do arquivo de parâmetros usado para rodar modelo animal no INTERGEN (Anexo 1)

2.3.3. Junta dados e soluções para o efeito ambiental e cria polinômios de Legendre de ordem *o*, especificada abaixo ($o = 1$ -> linear, intercepto e inclinação, $o = 2$ -> quadrático, etc.). Note que se a regressão aleatória é linear ($o = 1$), como tipicamente em modelos de normas de reação, não há necessidade de utilizar polinômios de Legendre.

2.3.4. Gera arquivo de dados para rodar modelo de normas de reação.

Para rodar o programa INTERGEN com um modelo de normas de reação, ver detalhes na Seção III a seguir e no exemplo apresentado no Anexo 2.

Seção III – Usando INTERGEN para análises de dados

3.1. Como rodar o programa:

Para rodar o programa, primeiro deve-se colocar os arquivos de parâmetros, de dados e de pedigree (e de reprodutores múltiplos, se for o caso) e o executável INTERGEN1.exe em um mesmo diretório. Depois abrir uma janela do prompt de comando do DOS, ir até o diretório onde foram colocados os arquivos e digitar: INTERGEN1 <enter>, o programa então irá solicitar o nome do arquivo de parâmetros, através da seguinte pergunta:
name of parameter file?
deve-se então digitar o *nome do arquivo de parâmetros* <enter>.

Essa resposta pode ser incluída em um arquivo texto (digamos *run*), p.ex., o arquivo *run* (formato ASCII) tem o seguinte conteúdo:
parameterfile

Assim podemos chamar o programa da seguinte forma:
C:\...caminho...\INTERGEN1 < run > out <enter>

Neste caso o INTERGEN obtém o nome do arquivo de parâmetros do arquivo *run* e toda a saída de tela é colocada no arquivo *out*.

O tempo de computação dependerá do número de observações e de equações no modelo e da velocidade de processamento do computador, podendo ser um limitante para a utilização do programa com modelos muito complexos envolvendo grande número de animais e parâmetros. Por exemplo, para analisar um conjunto de dados com aproximadamente 63.000 registros e 195.000 equações no modelo, o programa roda aproximadamente 10 ciclos por minuto em um PC com processador Intel Pentium IV 3.2 Ghz e 2GB de memória RAM, levando por volta de uma semana para concluir 100.000 ciclos.

IMPORTANTE: Muitos falhas ("crashes") do programa podem estar associados a erros na especificação do número de níveis dos efeitos no modelo e na especificação das posições dos efeitos, dentro do arquivo de parâmetros. Muitas vezes esses erros vêm associados com a palavra "hash" ou "ACCESS VIOLATION".

3.2. Como continuar uma cadeia já iniciada:

Em inferência MCMC, é comum necessitar alongar uma cadeia já finalizada para aumentar o conteúdo de informação ou após ter uma cadeia interrompida por falta de energia ou pela necessidade de desligar o computador. Nestes casos, o INTERGEN através da seção RESTART: Y/N? [CYCLE_TO_RESTART] oferece a oportunidade de continuar a cadeia por mais ciclos, sem ter que passar por novo período de aquecimento ou perder os ciclos de uma cadeia interrompida.

Para tal, o programa, cada vez que grava os arquivos de saída, grava também a última amostra de todos os parâmetros nos arquivos solutions, varcomp, structural_r e robustness_w (ver seção IV para detalhes dos arquivos).

Para continuar/reiniciar a cadeia, o usuário deve observar na primeira linha de um dos arquivos acima qual foi o último ciclo gravado, por exemplo, no *varcomp*:

Variance components after: 44000 rounds and burn-in of 10000.

Neste caso o último ciclo foi o 44.000 e, portanto, a cadeia deve ser reiniciada no 44.001, alterando a seção correspondente no arquivo de parâmetros da seguinte forma:
RESTART: Y/N? [CYCLE_TO_RESTART]
y 44001

Além disso, existe um passo não automático muito importante de ser completado pelo usuário. Uma vez que nos arquivos *solutionsam, varcompsam,* e *loglike_rnd* as amostras são salvas com maior freqüência que nos outros mencionados acima, o usuário deverá abrir manualmente com um editor de textos esses arquivos, ir até o final do arquivo e apagar todas as amostras salvas após o último ciclo gravado nos arquivos *solutions*, *varcomp, structural_r* e *robustness_w* (44.000 no exemplo acima), salvando os arquivos com os nomes originais sem colocar nenhuma extensão.

Seção IV – Inferência a posteriori a partir das saídas do programa

O programa INTERGEN utiliza o amostrador de Gibbs e Algoritmo de Metropolis-Hastings (SORENSEn; GIANOLA, 2002) para obter amostras dos parâmetros definidos no modelo, produzindo os seguintes arquivos de saída:

Solutions: Média e desvio padrão a posteriori dos efeitos no modelo e última amostra para continuar a cadeia

Solutionsam: Amostras dos efeitos no modelo, onde na seção EFFECTS foi escolhida a opção SAVE_SAMPLES = **y**

Varcomp: Média e desvio padrão a posteriori dos componentes de variância após o período de aquecimento e última amostra para recomeçar cadeia. Os efeitos dos RANDOM_GROUP aparecem primeiro e na mesma ordem que são especificados no arquivo parameterfile e por último aparecem os componentes de (co)variância residual.

Varcompsam: Amostras dos componentes de variância e, se for o caso, efeitos do modelo estrutural heteroscedástico. Os componentes de variância aparecem na mesma ordem do **varcomp** e das matrizes, que por serem simétricas, são listados somente elementos da diagonal inferior, da seguinte forma: v11, v12, v22, v13, v23, v33... Se for modelo estrutural robusto, são salvas após a variância residual amostras dos parâmetros de heteroscedasticidade dos fatores multiplicativos "aleatórios" e das soluções dos fatores multiplicativos "fixos" dos efeitos estruturais para variância residual da seção **STRUCTURAL_EFFECTS:** onde a opção **SAVE_SAMPLES?** = **y** e na mesma ordem que são especificados no parameterfile. Por último aparece a amostra do parâmetro de robustez.

structural_r: Média e desvio padrão *a posteriori* dos parâmetros de heteroscedasticidade (alpha) e soluções dos efeitos estruturas na variância residual e última amostra para recomeçar cadeia, lista-dos na mesma ordem que são especificados no parameterfile.

robustness_w: Média e desvio padrão *a posteriori* do parâmetro de robustez (alpha) e das variáveis de ponderação do modelo robusto e última amostra para recomeçar cadeia, para cada registro, na mesma ordem do arquivo de dados.

Loglike_rnd: Informações para critérios de escolha de modelo e medidas de ajuste, globais para todas as observações em cada ciclo MCMC.

Loglike_obs: Informações para critérios de escolha de modelo e medidas de ajuste, individuais para cada observação através de todos os ciclos MCMC.

Mh_dbeliefchg: Sintonia da variância das distribuições proposta do algoritmo de MH (se for o caso)

Mh_acceptance: Taxa de aceitação de valores propostos no algoritmo de MH (se for o caso)

4.1. Análise de convergência e variância de Monte Carlo

A análise de convergência é um processo fundamental na inferência bayesiana baseada em MCMC. Por ser um processo interativo de amostragem, as amostras seqüências da cadeia são correlacionadas e algum tempo é necessário para que a cadeia "esqueça" os valores iniciais e atinja a distribuição equilíbrio, da qual as amostras serão usadas para inferência. A velocidade que se da a convergência depende da complexidade do modelo, da correlação a posteriori dos parâmetros e da autocorrelação entre amostras sucessivas.

Apesar da sua importância, não existe consenso nem método automático para avaliação da convência de cadeias MCMC, essencialmente por que a densidade a posteriori da distribuição alvo é, em geral, desconhecida. Existem, entretanto, diversos métodos que podem ser usados como indicativos de provável convergência (COWLES; CARLIN, 1996; BROOKS; ROBERTS, 1998) e o uso concomitante de vários desses procedimentos é recomendável. Na prática, para as análises típicas em genética e melhoramento animal tem-se adotado o seguinte procedimento (baseado em Geyer (1992)):

A. Rodar uma cadeia "longa" dependendo da complexidade do modelo, algo entre 100.000 e 200.000 ciclos após um período de descarte/ aquecimento ("burn in") de 10% da cadeia e salvando a cada 10 ou 20 amostras ("thinning").

B. Fazer gráficos de traço dos valores das amostras para todos os componentes de variância versus o ciclo, incluindo a média da segunda metade da cadeia como referência no gráfico. Neste gráfico verificar que a cadeia tenha cruzado a linha da média pelo menos duas vezes no período de descarte e que após esse período as amostras variem sobre essa linha da média, sem tendência.

C. Calcular e observar as correlações de Pearson entre as amostras dos diferentes componentes de variância e autocorrelação de diferentes atrasos (p.ex. lag=1, 10 e 100) entre amostras de um mesmo componente de variância.

Os seguintes critérios são indícios de problemas de convergência:

- Correlações extremas (>0,9 ou <-0,9) entre diferentes parâmetros;

- Autocorrelações próximas a 1 em atrasos 1 e que não decresçam significativamente (para próximo a zero) quando o atraso é maior (p.ex. 100 ou até mesmo 1000);

- Gráficos de traço que apresentam um movimento muito lento ("slow mixing") da cadeia ao redor da média (gráficos em ondas indicam movimento lento enquanto ziguezagues indicam boa mistura da cadeia) ou apresentem tendência (não circulem ao redor da média) para algum parâmetro.

Um exemplo de aplicação do procedimento acima para análise de convergência de componentes de variância é apresentado no Anexo 7 (Rotina Pós-INTERGEN1.sas).

Caso sejam observados problemas de convergência, deve-se verificar a especifi-cação do modelo e ver se alguns dos parâmetros que apresentam alta correlação a *posteriori* podem ser suprimidos. Deve-se também rodar cadeias adicionais com outros valores iniciais e/ou outras sementes para o gerador de números aleató-rios. Sendo que uma possível estratégia é:

A. Rodar pelo menos duas outras cadeias de mesmo comprimento da já rodada uma iniciando com valores extremos para cima da média observada dos parâmetros e outra para baixo;

B. Fazer o gráfico de traço concomitante das três cadeias que iniciaram com valores distintos e verificar a convergência das mesmas para o mesmo espaço amostral, após o período de descarte;

C. Calcular o fator quantitativo de Gelman e Rubin (1992), que compara a variação dentro e entre cadeias. Se for próxima a 1 (não maior que 1,3) é indicativo de convergência.

Ainda uma terceira estratégia, válida somente para cadeias exclusivas pelo amostrador de Gibbs, sem passos de Metropolis, é o método do emparelhamento de cadeias de Johnson (1996), que consiste em iniciar duas cadeias com a mesma semente para o gerador de números aleatórios, mas com diferentes valores iniciais. Neste caso, a convergência é indicada pelo perfeito emparelhamento das cadeias após o período de aquecimento.

4.2. Inferência a partir da distribuição marginal a posteriori dos parâmetros

A partir das amostras salvas é possível reconstituir a distribuição marginal a posteriori de todos os parâmetros do modelo e de qualquer função dos mesmos que seja de interesse (p.ex. herdabilidades, correlações, diferenças entre médias de grupos, etc.). Uma ferramenta útil para este fim é o Procedimento KDE (kernel density estimator) do SAS (SAS INSTITUTE, 2004), que usando métodos não paramétricos estima a função densidade dos parâmetros ou funções, permitindo obter também médias, medianas, modas, variância, desvios padrão, percentis e intervalos de probabilidade a posteriori.

Outro aspecto importante na inferência a posteriori é a estimação da variância de Monte Carlo para os estimadores, que indica quanto que os valores obtidos para essas estatísticas podem variar em função da dimensão finita da cadeia. Pode-se também calcular o conteúdo de informação da cadeia (para todos os componentes de variância e demais parâmetros), que indica a quantas amostras independentes a cadeia se equivale. Caso não se identifique problema de convergência pelos critérios acima, mas for observado baixo número efetivo de amostras para algum parâmetro, deve se aumentar o número de amostras, reiniciando a cadeia ou rodando outras cadeias.

No Anexo 7 é apresentada um conjunto de macros do SAS (SAS INSTITUTE, 2004), (Pos_INTERGEN1.sas) para análise de convergência

e inferência a *posteriori*, usando PROC KDE e PROC ARIMA do SAS (SAS INSTITUTE, 2004) e critérios baseado em Geyer (1992) e Sorensen et al. (1995) para calcular respectivamente a variância de Monte Carlo e o número efetivo de amostras, conforme descrito a seguir.

Considere a seqüência de amostras da cadeia $X_1, X_2, \ldots, X_m$ onde m é o número de ciclos e X_i é a amostra no ciclo i. A auto-covariância temporal da seqüência é estimada por:

$$\hat{\gamma}_m(t) = \frac{1}{m} \sum_{i=1}^{i=m-t} (X_i - \hat{\mu}_m)(X_{i+t} - \hat{\mu}_m)$$

onde

$$\hat{\mu}_m = \frac{1}{m} \sum_{i=1}^{i=m} X_i$$

é a média amostral para a cadeia e t é o atraso. A autocorrelação de atraso (t) é estimada por:

$$\frac{\hat{\gamma}_m(t)}{\hat{\gamma}_m(0)}$$

Um estimador da variância da média amostral proposto por Geyer (1992), baseado em series temporais, chamado estimador da seqüência positiva inicial. Dada a função

$$\hat{\Gamma}_m(t) = \hat{\gamma}_m(2t) + \hat{\gamma}_m(2t+1), \qquad t = 0,1,\ldots,$$

o estimador é definido como:

$$m(\mathrm{var}(\hat{\mu}_m)) = \hat{\gamma}_m(0) + 2\sum_{i=1}^{2t+1} \hat{\gamma}_m(i) = -\hat{\gamma}_m(0) + 2\sum_{i=0}^{t} \hat{\Gamma}_m(i)$$

sendo *t* o maior número inteiro que satisfaça $\hat{\Gamma}_m(i) > 0, \quad i = 0,1,...,t$.
O número efetivo de amostras independentes (SORENSEN et al., 1995) é computado por:

$$\hat{\psi}_m = \frac{\hat{\gamma}_m(0)}{\text{var}(\hat{\mu}_m)}$$

O método de *batching* também pode ser usado para estimar a variância de Monte Carlo e o tamanho amostral (SORENSEN; GIANOLA, 2002)

4.3. Critérios de escolha de modelo

Outro ponto fundamental nas análises é a escolha do modelo a ser utilizado, especialmente quando é preciso decidir entre duas ou mais alternativas. Na inferência Bayesiana a escolha do modelo é, em geral, baseada na distribuição marginal dos registros, buscando o modelo que melhor prediz o que foi de fato observado nos dados. Normalmente avalia-se o desempenho global do modelo no ajuste de todos os registros, entretanto em alguns casos pode ser interessante avaliar o modelo com respeito a um subconjunto de observações, de interesse diferenciado.

Três critérios podem ser prontamente utilizados a partir dos arquivos de saída do Programa INTERGEN, os quais têm sua implementação facilitada pela rotina Pos_INTERGEN2.sas no Anexo 8:

4.3.1. Critério de Informação da Deviance (DIC):

O DIC (SPIEGELHALTER et al., 2002) é um meio de comparação de modelos que segue a proposição de Dempster (1997), o qual sugere que comparações entre modelos sejam baseadas de distribuição a posteriori da deviance de cada modelo. O DIC é composto por uma medida de ajuste global – a média a posteriori da deviance - e uma penalização por complexidade do modelo (número efetivo de parâmetros, pD) – a diferença entre a média a posteriori da deviance e a deviance baseada na média a posteriori dos parâmetros do modelo.

A *deviance* do modelo *i* pode ser definida como:

$$D(\boldsymbol{\theta})_i = -2\log p(\mathbf{y} \mid \boldsymbol{\theta}, M_i)$$

Um estimador de Monte Carlo é obtido por:

$$\bar{D}(\boldsymbol{\theta})_i = \tfrac{1}{m}\sum_{j=1}^{m} -2\log p(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i)$$

A complexidade do modelo *i* como número efetivo de parâmetros é dada por:

$$p_{Di} = \bar{D}(\boldsymbol{\theta})_i - D(\bar{\boldsymbol{\theta}})_i$$

onde $D(\bar{\boldsymbol{\theta}})_i = -2\log p(\mathbf{y} \mid \bar{\boldsymbol{\theta}}, M_i)$ e $\bar{\boldsymbol{\theta}}$ é a media *a posteriori* dos parâmetros. Finalmente o *DIC* é calculado por:

$$DIC_i = \bar{D}(\boldsymbol{\theta})_i + p_{Di}$$

Obtem-se $\log p(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i) = \sum_{k=1}^{n} \log p(y_k \mid \boldsymbol{\theta}^{(j)}, M_i)$ e ao final de *m* ciclos avalia-se $\bar{\boldsymbol{\theta}} = \tfrac{1}{m}\sum_{j=1}^{m} \boldsymbol{\theta}^{(j)}$ para obter,

$D(\bar{\boldsymbol{\theta}})_i = \sum_{k=1}^{n} -2\log p(y_k \mid \bar{\boldsymbol{\theta}}, M_i)$, a partir da soma dos valores da 4ª coluna do arquivo loglike_obs, e dado que:

$$DIC_i = 2\bar{D}(\boldsymbol{\theta})_i - D(\bar{\boldsymbol{\theta}})_i$$

Menores valores de *DIC* indicam melhor ajuste do modelo.

4.3.2. Pseudo Fator de Bayes (PBF):

O PBF é calculado a partir da ordenada preditiva condicional (conditional predictive ordinate - CPO), conforme descrito por Gelfand (1996). As CPO's são densidades de validação cruzada $p(y_k \mid \mathbf{y}_{(k)})$, as quais su-

gerem quais valores de y_k são prováveis quando o modelo é ajustado a todas as observações exceto y_k (um conjunto representado por $y_{(k)}$). A CPO proporciona uma medida de ajuste para cada observação individualmente e comparações entre modelos são feitas por razões de CPO's:

$$C_k = \frac{p\left(y_k \mid \mathbf{y}_{(k)}, M_1\right)}{p\left(y_k \mid \mathbf{y}_{(k)}, M_2\right)}$$

Adicionalmente, uma medida global de ajuste é dada pelo PBF

$$PBF = \prod_{k=1}^{n} C_k = \prod_{k=1}^{n} \frac{p\left(y_k \mid \mathbf{y}_{(k)}, M_1\right)}{p\left(y_k \mid \mathbf{y}_{(k)}, M_2\right)}$$

Desde que y_k sejam condicionalmente independentes dado $\boldsymbol{\theta}$, o CPO pode ser estimado por Monte Carlo da seguinte forma:

$$CPO_k = p\left(y_k \mid \mathbf{y}_{(k)}, M_i\right) = \frac{1}{\frac{1}{m}\sum_{j=1}^{m} p^{-1}\left(y_k \mid \boldsymbol{\theta}^{(j)}, M_i\right)}$$

O programa INTERGEN avalia a cada ciclo $p\left(y_k \mid \boldsymbol{\theta}^{(j)}, M_i\right)$, mantendo uma soma corrente desses valores para no final calcular a média harmônica acima, que é salva para cada observação na 2ª coluna do arquivo loglike_obs. Maiores valores de CPO indicam melhor ajuste.

Um estimador estável de PBF é dado por:

$$PBF = \exp\left(\sum_{k=1}^{n} \log \hat{p}\left(y_k \mid \mathbf{y}_{(k)}, M_1\right) - \sum_{k=1}^{n} \log \hat{p}\left(y_k \mid \mathbf{y}_{(k)}, M_2\right)\right)$$

Finalmente o estimador da deviance do modelo baseada em CPO é obtido por

$$-2\sum_{k=1}^{n} \log \hat{p}\left(y_k \mid \mathbf{y}_{(k)}, M_1\right)$$

4.3.3. Fatores de Bayes (BF):

O fator de Bayes é uma medida global de ajuste data pela razão entre as distribuições marginais dos dados $p(\mathbf{y} \mid M_i)$ sob dois modelos diferentes (M_i, i = 1,2), dado que

$$BF = \frac{p(\mathbf{y} \mid M_1)}{p(\mathbf{y} \mid M_2)}$$

Pode ser demonstrado que $p(\mathbf{y} \mid M_i) = E^{-1}\left[p^{-1}(\mathbf{y} \mid \boldsymbol{\theta}, M_i)\right]$ e um estimador de Monte Carlo foi proposto por Newton and Raftery (1994).

$$\hat{p}(\mathbf{y} \mid M_i) = \frac{1}{\frac{1}{m}\sum_{j=1}^{m} p^{-1}\left(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i\right)}$$

onde m é o número do ciclos de Gibbs e $\boldsymbol{\theta}^{(j)}$ é a amostra da distribuição a *posteriori* dos parâmetros no ciclo *j*. Para evitar erros de precisão ao calcular BF a estratégia adotada é a seguinte:

$$f_i = \frac{1}{m}\sum_{j=1}^{m} \exp\left(-\log p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i\right) - c\right)\exp c$$

onde c é o maior valor de $-\log p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i\right)$ e tomando-se o logaritmo

$$\log f_i = \log\left[\frac{1}{m}\sum_{j=1}^{m} \exp\left(-\log p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i\right) - c\right)\right] + c$$

finalmente

$$BF = \exp\left(-\log f_1 + \log f_2\right)$$

A cada ciclo o Programa INTERGEN salva no arquivo loglike_rnd $\log p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i\right) = \sum_{k=1}^{n} \log p\left(y_k \mid \boldsymbol{\theta}^{(j)}, M_i\right)$ e no final dos m ciclos o *BF* pode ser obtido como descrito acima, sendo que valores maiores que um suportam M_1, enquanto valores menores que um suportam M_2. Não há necessidade de os modelos serem aninhados como no teste da razão da verossimilhança, mas é preciso que todas as especificações de priores

sejam próprias para que as comparações sejam válidas. Finalmente a deviance do modelo *i* baseado no estimador dos Fatores de Bayes pode ser obtida por -2 log f_1 e neste caso menores valores indicam um melhor ajuste, isto é, um menor desvio do modelo hipotético de ajuste perfeito.

Termo de isenção de responsabilidade: O Programa INTERGEN, bem como as rotinas providas neste manual, foram desenvolvidos para pesquisa científica em genética animal e são disponibilizados sem custo, para este fim. O programa foi testado e validado pelo autor em diversas análises, entretanto, como nem todas as possibilidades foram testadas, **o autor não pode ser responsabilizado por eventuais erros ("bugs") que o programa possa conter e pelo uso de seus resultados.** O código Fortran 90/95 está disponível por solicitação e comentários, críticas e sugestões são bem vindas.

Referências

BROOKS, S. P.; ROBERTS, G. O. Convergence assessment techniques for Markov chain Monte Carlo. **Statistics and Computing**, v. 8, n. 4, p.319-335, Dec. 1998.

CARDOSO, F. F.; ROSA, G. J. M.; TEMPELMAN, R. J. Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. **Journal of Animal Science**, Savoy, v. 83, n. 8, p.1766-1779, Aug. 2005.

COWLES, M. K.;. CARLIN, B. P.. Markov chain Monte Carlo convergence diagnostics: a comparative review. **Journal of the American Statistical Association**, New York, v. 91, n. 434,, p. 883-904, Jun. 1996.

DEMPSTER, A. P. The direct use of likelihood for significance testing (reprinted from Memoirs No. 1, Proceedings of Conference on Foundational Questions in Statistical Inference, Aarhus, Denmark, 7-22 May 1973, pp 335-54). **Statistics and Computing**, London, v. 7, n. 4, p. 247-252, Dec. 1997.

GELFAND, A. E. Model determination using sampling-based methods. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. (Ed.). **Markov chain Monte Carlo in practice.** London: Champman & Hall, 1996. p. 145-161.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-472, Nov. 1992.

GEYER, C. J. Practical Markov chain Monte Carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 473-511, Nov. 1992.

HENDERSON, C. R. **Aplications of linear models in animal breeding.** Guelph: University of Guelph. 1984. 462 p.

JOHNSON, V. E. Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. **Journal of the American Statistical Association**, New York, v. 91, n. 433, p.154-166, Mar. 1996.

LO, L. L.; FERNANDO, R. L.; GROSSMAN, M. Covariance between relatives in multibreed populations - additive-model. **Theoretical and Applied Genetics**, New York, v. 87, n. 4, p. 423-430, Dec. 1993.

MISZTAL, I. BLUPF90 - a flexible mixed model program in Fortran 90. Disponível em: <http://nce.ads.uga.edu/~ignacy/newprograms.html>. Acesso em: mar. 2007.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7.,2002, Montpellier. **Anais**... Montpellier: INRA, 2002. Session 28. 1 CD-ROM.

NEWTON, M. A. E A. E. RAFTERY. Approximate Bayesian-Inference With The Weighted Likelihood Bootstrap. **Journal of the Royal Statistical Society Series B-Methodological**, Oxford, v.56, n.1, p.3-48. 1994.

ROSO, V. M.; SCHENKEL, F. S. AMC – a computer program to assess the degree of connectedness among contemporary groups. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Anais**... Belo Horizonte: Instituto Prociência, 2006. Session 27-26. 1 CD-ROM.

SAS INSTITUTE. **SAS OnlineDoc 9.1.3.** Cary, NC, 2004. Disponível em: <http://support.sas.com/onlinedoc/913/docMainpage.jsp>. Acesso em: mar. 2007.

SORENSEN, D. A.; ANDERSEN, S.; GIANOLA, D.; KORSGAARD, I. Bayesian-inference in threshold models using Gibbs sampling. **Genetics Selection Evolution**, Paris, v. 27, n. 3, p. 229-249, 1995.

SORENSEN, D. A.; GIANOLA, D. **Likelihood, Bayesian and MCMC methods in quantitative genetics**. New York: Springer-Verlag, 2002. 740 p.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. R.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society Series B-Statistical Methodology**, Oxford, v. 64, n. 4, p. 583-616, Oct. 2002.

SU, G.; MADSEN, P.; LUND, M. S.; SORENSEN, D.; KORSGAARD, I. R.; JENSEN, J. Bayesian analysis of the linear reaction norm model with unknown covariates. **Journal Of Animal Science**, Champaign, v. 84, n. 7, p.1651-1657, Jul. 2006.

Anexos

Anexo 1. Exemplo de modelo animal

O arquivo de parâmetros abaixo consiste de um exemplo para rodar um típico modelo animal com resíduos normais e homogêneos, onde grupos contemporâneos são aleatórios.

O modelo é o seguinte:

y = animal + constante + grupo de contemporâneos + idade da vaca linear + idade da vaca quadrático + idade do animal linear + idade do animal quadrático + erro

```
# Angus Pos-desmama Modelo Animal
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
100000 10000 10
SEED
123
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
Angusgpd.dat 63098
NUMBER_OF_TRAITS
1
```

```
NUMBER_OF_EFFECTS
7
OBSERVATION(S)
9
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT  [EFFECT NESTED]
1 95896 cross n       #animal
2 1 cov n             #media_geral_coluna_de_1s
3 2482 cross n        #gc_aleatorio
5 1 cov n             #idv
6 1 cov n             #idv2
7 1 cov n             #ids
8 1 cov n             #ids2
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 1
RESIDUAL_PRIOR_(CO)VARIANCES
400
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
Angusgpd.ped 95896 1 0
(CO)VARIANCES
200
RANDOM_GROUP
3
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
diagonal 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]

(CO)VARIANCES
2800
```

Anexo 2. Exemplos de modelos de normas de reação

A2.1. Exemplo de modelo de normas de reação em dois passos (usando soluções ambientais de análise prévia)

Esse arquivo de parâmetros consiste de um exemplo para rodar um modelo de normas de reação para estudos de interação genótipo-ambiente com resíduos normais e homogêneos. Estimativas dos efeitos ambientais previamente estimados, usando o modelo do Anexo 1, se encontram na coluna 4 do arquivo de dados após rodar o Pre_Intergen3.sas (refere-se onde grupos de contemporâneos). Entretanto o efeito de grupo de contemporâneos é estimado novamente com efeito aleatórios, permitindo um ajuste para esse efeito na presença de interação genótipo-ambiente.

O modelo é o seguinte:

y = constante + grupo de contemporâneos + animal intercepto + animal linear + idade da vaca linear + idade da vaca quadrático + idade do animal linear + idade do animal quadrático + erro

```
# Angus Pos-desmama Modelo de Normas de Reação – dois passos
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
200000 20000 100
SEED
123
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
Angusgpd_iga.dat 63098
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
8
OBSERVATION(S)
12
WEIGHT(S)
```

```
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES [EFFECT NESTED]
2 1 cov n            #media_geral_coluna_de_1s
3 2482 cross n        #gc_aleatorio
2 95896 cov n 1       #animal_intercepto
4 95896 cov n 1       #animal_linear
8 1 cov n            #idv
9 1 cov n            #idv2
10 1 cov n            #ids
11 1 cov n            #ids2
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 1
RESIDUAL_PRIOR_(CO)VARIANCES
600
RANDOM_GROUP
2
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
diagonal 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
(CO)VARIANCES
2800
RANDOM_GROUP
3 4
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal 2
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
Angusgpd.ped 95896 1 0
(CO)VARIANCES
100 10
10  5
```

A2.2. Exemplo de modelo de normas de reação em um único passo

Esse arquivo de parâmetros também consiste de um exemplo para rodar um modelo de normas de reação para estudos de interação genótipo-ambiente com resíduos normais e homogêneos, entretanto as estimativas dos efeitos ambientais são obtidas conjuntamente com a norma de reação em uma única análise de acordo com Su et al. (2006), sem necessidade de usar resultados de análise prévia.

Note que apesar de o modelo ser o mesmo acima (A2.1), o processo de estimação é diferente e simultânea para GC e norma de reação. No arquivo de parâmetros abaixo o efeito de GC na segunda linha da seção *EFFECTS* é do tipo covariável desconhecida para modelo de norma de reação (**unknowncov**), pois as soluções para este efeito são usadas como covariável para obter a norma de reação. Este efeito de norma de reação aparece na quarta linha da seção EFFECTS (tipo **nrnom**) e, neste caso de análise em um passo, faz-se referência a coluna 3 do arquivo de dados onde estão as identificações dos GC e não mais a coluna 4 dos dados (que contém as soluções de GC do modelo animal rodado previamente) como no exemplo acima (A2.1).

```
# Angus Pos-desmama Modelo de Normas de Reação em um passo
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
200000 20000 100
SEED
123
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
Angusgpd_iga.dat 63098
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
8
OBSERVATION(S)
12
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES [EFFECT NESTED]
2 1 cov n
3 2482 unknowncov n
2 95896 cov n 1
3 95896 rnorm n 1
8 1 cov n
9 1 cov n
10 1 cov n
11 1 cov n
```

```
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 1
RESIDUAL_PRIOR_(CO)VARIANCES
600
RANDOM_GROUP
2
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
diagonal 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]

(CO)VARIANCES
2800
RANDOM_GROUP
3 4
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal 2
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
Angusgpd.ped 95896 1 0
(CO)VARIANCES
100 10
10  5
```

Anexo 3. Modelo Multirracial

```
# multiple-breed Gaussian heterogeneous structural model
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
200000 20000 100
SEED
1234
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
mbham_dat1  4000
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
6
OBSERVATION(S)
8
```

```
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT  [EFFECT NESTED]
1   4000 cross n  #efeito-animal-multirracial
3 1 cov y         #media-geral
4 1 cov y         #efeito-sexo-(como-dummy-variable)
5 1 cov y         #proporcao-racial
6 1 cov y         #heterozigose
7   200 cross n   #GC-aleatorio
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
structural 8
METROPOLIS_STEP_OF_STRUCTURAL_EFFECTS:
ROUNDS_WITHIN_CYCLE TUNING_SKIP
1 10
NUMBER_OF_STRUCTURAL_EFFECTS
4
STRUCTURAL_EFFECTS: LINE_FROM_EFFECTS_SECTION
SAVE_SAMPLES?
3 y
4 y
5 y
6 y
RESIDUAL_PRIOR_(CO)VARIANCES
200
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_mb 8
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
mbham_pedt1  4000 2 0
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS]
0
METROPOLIS_STEP_OF_MULTIBREED_(CO)VARIANCES:
ROUNDS_WITHIN_CYCLE TUNING_SKIP
1 10
(CO)VARIANCES
49      28
28      91
```

RANDOM_GROUP
6
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
diagonal 8
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS

(CO)VARIANCES
200

Anexo 4 – Pré_INTERGEN1.sas:

```
* Pre_INTERGEN1.sas - Rotina para preparação e consistência de dados
para o programa INTERGEN,
* até análise de conectabilidade (Programa AMC, Roso & Schenkel,
2006).
* Por Fernando F. Cardoso 2006.
* Dados Angus do PROMEBO pós-desmama arquivo extensão .INS ;
options nodate;
title1 'Promebo Angus';
* 1. Leitura do arquivo original de dados;
data abg0ins;
/* Modifique o caminho conforme a localização do arquivo*/
  infile 'C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\ABG0.INS' lrecl=91;
  input
/*
CAMPO                  TAMANHO      INÍCIO       DESCRIÇÃO
*/
TERN $ 1-6 /*             A6           1           Identificação do
produto */
NA 7-8 /*                 I2           7            Ano de nascimento
produto */
C $ 9-16 /*               A8           9          Criador-Raça-Rebanho
(A4,A3,A1) */
ANOP 17-18 /*             I2          17       Ano da produção (safra)
*/
NEST $ 19 /*              A1          19       Estação de produção */
CSD $ 20 /*               A1           20   Código do sexo desmama
*/
```

CA 21-22 /*	I2	21	Grupo de manejo desmama
(Código alimentar) */			
JA 23-25 /*	I3	23	Data da pesagem de
desmama */			
CSP $ 26 /*	A1	26	Código do sexo pós-
desmama (final) */			
CAP 27-28 /*	I2	27	Código de manejo pós-
desmama */			
JAP 29-31 /*	I3	29	Data da pesagem final */
TOURO0 $ 32-37 /*A6,2X		32	Touro pai */
ANT 38-39			
VACA0 $ 40-45 /*	A6	40	Mãe */
ANV 46-47 /*	I2	46	Ano de nascimento da mãe
*			
GPDA 48-55 /*	F8.3	48	Ganho de peso pós-
desmama ajustado */			
VAECS 56-57 /*	I2	56	Valor aritmético do escore
conformação final */			
ECS 58-59 /*	I2	58	Escore de conformação
final */			
SIES $ 60 /*	A1	60	Sinal do escore de
conformação final */			
EDS 61 /*	I1	61	Estado ou condição
corporal final */			
GORF 62 /*	I1	62	Escore final para
precocidade de terminação */			
MUSF 63 /*	I1	63	Escore final para
musculosidade */			
TAMF 64 /*	I1	64	Escore final para tamanho
*/			
ANKF $ 65-66 /*	A2	65	Escores Ankony (U e R)
finais -restantes */			
GNDA 67-74 /*	F8.3	67	Ganho de peso
nascimento-desmama ajustado */			
IDV 75-76 /*	I2	75	Idade da vaca */
DJN 77-79 /*	I3	77	Data de nascimento juliana
produto */			
IDD 80-82 /*	I3	80	Idade do terneiro a
desmama */			
DTP 83-85 /*	I3	83	Dias em teste pós-
desmama */			

```
DTP 83-85 /*            I3              83      Dias em teste pós-
desmama */
PD 86-88 /*             I3                      86              Peso real a
desmama */
PS 89-91 /*             I3                      89              Peso real
final */
;
* 2. Cria grupos de contemporâneos;
* Criador-ano-estação-código manejo e data da desmama-código manejo
e data do sobreano;
GC=compress(C||ANOP||NEST||CA||JA||CAP||JAP||CSP,' ');
GC_s=compress(C||ANOP||NEST||CA||JA||CAP||JAP,' ');
GC_s_ja=compress(C||ANOP||NEST||CA||CAP,' ');
RAE=compress(C||ANOP||NEST,' ');
* 3. Cria identificação única de produtos e vacas e hbb touros (touro);
animal=compress(TERN||NA||C,' ');
vaca=compress(VACA0||ANV||C,' ');
* 4. Cria variáveis;
GPD = PS-PD; * ganho pós desmama;
GDPD = GPD/DTP; * ganho diário pós desmama;
* ano nascimento do animal, touro e vaca;
if na<50 then anp=na+2000;
  else anp=na+1900;
if anv<50 then anv=anv+2000;
  else anv=anv+1900;
if ant<50 then ant=ant+2000;
  else ant=ant+1900;
if IDV=0 then delete;
IDV2=IDV**2; * idade da vaca ao quadrado;
if CSP^='1' & CSP^='2' then delete;
IDS=IDD+DTP; * idade ao sobreano = idade a desmama + dias em
teste pós-desmama;
IDS2=IDS**2;
DTP2=DTP**2;
* Identifica animais filhos de reprodutores múltiplos (rm=1)
(Particular do Promebo);
if (substr(touro0, 1, 2)='RM') then rm=1;
  else if (substr(touro0, 2, 2)='RM') then rm=1;
  else if (substr(touro0, 3, 2)='RM') then rm=1;
  else if (substr(touro0, 4, 2)='RM') then rm=1;
  else if (substr(touro0, 5, 2)='RM') then rm=1;
```

```
  else if (substr(touro0, 5, 2)='RM') then rm=1;
  else if (substr(touro0, 4, 1)='R' & substr(touro0, 5, 1)=' ') then
rm=1;
  else if (substr(touro0, 4, 1)='D' & substr(touro0, 5, 1)=' ') then
rm=1;
  else if (substr(touro0, 5, 1)='R' & substr(touro0, 6, 1)=' ') then
rm=1;
  else if (substr(touro0, 5, 1)='D' & substr(touro0, 6, 1)=' ') then
rm=1;
  else if (substr(touro0, 6, 1)='R') then rm=1;
  else if (substr(touro0, 6, 1)='D') then rm=1;
  else rm=0;
* Consistência o HBB de touros-pais com o cadastro de registro da ANC
(Particular do Angus Promebo/ANC);
if rm=0 then
  if (substr(touro0, 1, 1)='C') then
          touro=compress('C0'||substr(touro0, 2, 5));
  else if (substr(touro0, 1, 2)='IA') then
          touro=compress('IA-'||substr(touro0, 3, 3));
  else if (touro0>99999 & touro0^=.) then
          touro=compress('C'||substr(touro0, 1, 6));
  else if (touro0<100000 & touro0^=.) then
          touro=compress('00'||substr(touro0, 1, 5));
  else touro=touro0;
else touro=touro0;
RUN;
* Elimina possiveis duplicatas, mantendo o primeiro registro das
duplicatas;
proc sort data=abg0ins nodupkey; by animal; run;
/* Análise do ganho pós-desmama */
title2 'Ganho Pós-Desmama';
* 5. Estatísticas descritivas;
proc freq data=abg0ins;
  tables Gc*CSP IDV CSP rm/
  nocol norow nocum nopercent;
run;
proc means data=abg0ins;
  var IDV DJN IDD DTP IDS PD PS GPD GDPD;
run;
proc corr data=abg0ins;
  var DJN IDD DTP IDS PD PS GPD GDPD;
```

```
  var DJN IDD DTP IDS PD PS GPD GDPD;
run;
proc univariate data=abg0ins plot;
  var DTP IDS;
run;
* 6. Consistência;
data gpd;
  set abg0ins;
  * Elimina registros extremos de IDS e DTP detectados no PROC
UNIVARIATE;
  if IDS > 330 & IDS < 730;
  if DTP > 125 & DTP < 525;
  if gpd^=.;
run;
* Calcular média e desvio padrão (DP) por grupo contemporâneo;
proc sort data=gpd; by gc; run;
proc means data=gpd mean std noprint;
  by gc;
  var GPD;
  output out=mediagc mean=mediagpd std=dpgpd;
run;
proc freq data=mediagc;
  tables _FREQ_ /norow nocol;
run;
* Calcula o DP padronizado do GPD em relação ao GC;
data gpd1;
  merge mediagc gpd; by gc;
  if dpgpd>0 then dpgpdi=abs((gpd-mediagpd)/dpgpd);
run;
*Elimina GC com 5 ou menos obs e animais com mais de 3,5 DP do GC;
data gpd1;
  set gpd1;
  if dpgpdi=. | dpgpdi > 3.5 | _FREQ_ <6 then delete;
run;
/*Elimina touros n filhos < 2 */
proc freq data=gpd1 noprint; tables touro
/out=countt norow nocol nocum nopercent;
run;
proc sort data=gpd1;by touro;run;
proc sort data=countt;by touro;run;
```

```
data gpd2; retain count;
  merge countt gpd1; by touro;
  tocount=count;
  if count<2 then delete;
  drop percent count;
run;
title3 'Estatísticas pós-consistência';
*Estatísticas descritivas após consistência;
proc freq data=gpd2;
  tables IDV CSP rm GC touro/
  nocol norow nocum nopercent;
run;
proc means data=gpd2;
  var IDV DJN IDD DTP IDS PD PS GPD GDPD;
run;
proc corr data=gpd2;
  var DJN IDD DTP IDS PD PS GPD GDPD;
run;
proc univariate data=gpd2 plot;
  var DTP IDS GPD;
run;
* 7. Teste de conectabilidade de GC baseado no número total de laços
genéticos (mínimo 10), baseado em um modelo animal, usando o
programa AMC (Roso & Schenkel, 2006)
    Para rodar esse programa é necessário criar um arquivo de entrada
  "infile.txt" com o formato Fortran FORMAT(3a10,21x,a7,a1,24x,a1)
  para as variáveis ID de animal, touro e vaca (Alfa-númericas com
  tamanho 10), 21 espaços brancos, GC (Alfa-númericas com tamanho
7),
  sexo (M,F,Alfa-númericas com tamanho 1), 24 espaços brancos e o
tipo
  de serviço/paternidade (I e C para pai conhecido e M para RMs ou pai
  desconhecido,Alfa-númericas com tamanho 1)
    Neste caso do arquivo do promebo, como as IDs tem formato maior
  que 10 é necessário renumerar os animais antes de rodar o programa;

/* Renumera animais e cria arquivo de pedigree */
data ped;
  set gpd2;
*  keep animal touro vaca gc csd rm;
```

```
run;
proc sort nodupkey; by animal;
run;
data touro;
  set ped(where=(rm=0));  /* le os touros*/
  animal=touro;
  keep animal;
run;
proc sort nodupkey; by animal;      /* elima duplicatas */
run;
data vaca;
  set ped;                    /* le as vacas */
  animal=vaca;
  keep animal;
run;
proc sort nodupkey; by animal;      /* elima duplicatas */
run;
data animal;
  set ped vaca touro;
  keep animal;
run;
proc sort nodupkey; by animal;      /* elima duplicatas */
run;
proc sort data=ped nodupkey; by animal;      /* elima duplicatas */
run;
/* renumera animais (inclusive touros e vacas) de 1 a n de animais */
data ped2;
  merge animal(in=temp) ped;
  by animal;
  if temp=1;
  ianimal+1;
*  keep animal touro vaca gc csd rm ianimal;
run;
/* Junta as identificação original e recodificada de touros*/
data touro;
  set ped2;
  touro=animal;
  keep touro ianimal;
run;
proc sort data=ped2; by touro; run;
```

```
proc sort data=touro nodupkey; by touro; run;
data ped2; /* renumera touros */
  merge ped2(in=temp) touro(in=temp1 rename=(ianimal = id));
  by touro;
  if temp=1;
  if temp1 =1 then itouro=id;
  drop id;
run;
/* Junta as identificação original e recodificada de vacas*/
data vaca;
  set ped2;
  vaca=animal;
  keep vaca ianimal;
run;
proc sort data=ped2; by vaca; run;
proc sort data=vaca nodupkey; by vaca; run;
data ped2; /* renumera vacas */
  merge ped2(in=temp) vaca(in=temp1 rename=(ianimal = id));
  by vaca;
  if temp=1;
  ivaca=id;
  if itouro =. then itouro =0;
  if ivaca =. then ivaca =0;
  drop id;
run;
proc sort data=ped2; by animal; run;
* Renumera grupos de contemporâneos de 1 a n;
proc sort data=ped2;  by gc; run;
data ped2;
  set ped2;
  by gc;
  if first.gc then igc+1;
run;
proc sort data=ped2; by ianimal; run;
* Gera arquivo de entrada infile.txt para o programa AMC;
data _null_;
  set ped2;
file 'C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\AMC\infile.txt';
  tsd='I';
```

```
gender='F';
  if rm^=. & csd^='' & gc^='';
  if csd=1 | csd=3 then gender='M';
  if rm=1 then tsd='M'; * Identifica registros de filhos de rm;
  put ianimal 3-10 0 itouro 13-20 0 ivaca 23-30 0 igc 52-58 0 gender
$59 tsd $84;
run;

* Grava arquivo de pedigree e dados antes do teste de conectabilidade;
data _null_;
  set ped2;
  file "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\Angusgpd_preamc.dat";
   * listar aqui o animal, touro, vaca, gc, ianimal, itouro, ivaca, igc, efeitos
fixos e aleatórios e variáveis resposta a serem incluidos no modelo;
  if rm^=. & csd^='' & gc^='';
  put  animal 1-26 touro 27-34 vaca 35-60 gc 61-130 /*importante
colocar as colunas para evitar erros de leitura das variaveis caracter
desse arquivo pelo programa Pre-INTERGEN2.sas*/
       ianimal itouro ivaca igc
             rm CSP IDV IDV2 IDS IDS2
             gpd;
run;
* DEVE-SE RODAR o Programa AMC (ROSO; SCHENKEL, 2006) e
prosseguir com a rotina do SAS (SAS INSTITUTE, 2004) no arquivo
Pre_Integen2.sas ;
```

Anexo 5 – Pré_INTERGEN2.sas:

```
* Pre_INTERGEN2.sas - Rotina para preparação e consistência de dados
para o programa INTERGEN,
* da análise de conectabilidade Programa AMC, Roso & Schenkel (2006)
a geração dos arquivos
* de dados para o modelo animal
* IMPORTANTE: requer que a rotina Pre_INTERGEN1.sas tenha sido
rodada.
* Por Fernando F. Cardoso 2006.;
options nodate;
title1 'Pre_INTERGEN2 - Promebo Angus';
* 1. Le arquivo de saida do Programa AMC (animal.amc) identificando os
```

```
arquipelagos e os animais conectados e desconectados;
data conect;
* fornecer o caminho onde está o arquivo;
  infile "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\AMC\animal.amc";
  input ianimal 1-11 itouro 12-21 ivaca 22-31 gender $ 32 arquip 33-34
conect $ 36;
run;
* Lista animais desconectados;
proc print data=conect;
  where arquip^=1 | conect^='C';
  var ianimal arquip conect;
run;
* Le arquivo de pedigree e dados antes do teste de conectabilidade;
data gpd2;
  infile "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\Angusgpd_preamc.dat";
   * listar aqui o animal, touro, vaca, ianimal, itouro, ivaca, efeitos fixos e
aleatórios e variáveis resposta a serem incluidos no modelo;
  input  animal $ 1-26 touro $ 27-34 vaca $ 35-60 gc $ 61-130
/*importante colocar as colunas para evitar erros de leitura das variaveis
caracter desse arquivo pelo programa Pre-INTERGEN2.sas*/
       ianimal itouro ivaca igc
             rm CSP IDV IDV2 IDS IDS2
             gpd;
run;

* Grava arquivo de pedigree e dados antes do teste de conectabilidade;
data gpd2;
  set ped2;
  if rm^=. & csd^='' & gc^='';
run;
* Junta os arquivos de dados e de conectabilidade, mantendo somente
os
  animais conectados no arquipelago principal (1);
proc sort data=gpd2; by ianimal; run;
proc sort data=conect; by ianimal; run;
data gpd3;
  merge gpd2 conect(in=temp);
  by ianimal;
```

```
if arquip=1 & conect='C';
run;
/* 2. Arquivo de Pedigree */
/* Renumera animais conectados e cria arquivo de pedigree */
data ped3;
  set gpd3;
  keep animal touro vaca rm;
run;
proc sort nodupkey; by animal;
run;
data touro;
  set ped3(where=(rm=0));  /* le os touros*/
  animal=touro;
  keep animal;
run;
proc sort nodupkey; by animal;      /* elima duplicatas */
run;
* OPCIONAL: consulta no sumário de touros da ANC - raça Angus,
foram
  obtidos os nome, o pai e nome do pai do touro para aumetar laços
genéticos;
PROC IMPORT OUT= WORK.sire
            DATAFILE= "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\Touros pais.xls"
           DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
RUN;
data sire2;
  set sire(rename=(nome=name hbb=animal));
  name=trim(name);
  animal=compress(animal,' ');
  if substr(name,1,5)^='XXXXX';
  if substr(name,1,2)^='  ';
run;
proc sort nodupkey; by animal;      /* elimina duplicatas */
run;
data sire3; /* junta touros no arquivo com nome a ano de nascimento */
  merge sire2(in=temp1) touro(in=temp);
  by animal;
  if temp=1;
```

```
 if temp1=0 then name=animal;
run;
data sire4; /* obtem os pais dos touros (avos) no sumário */
  set sire3;
  name=trim(nome_pai);
  animal=compress(hbb_pai,' ');
  if animal^=' ';
  keep animal name;
run;
/*Identifica n filhos dos avos paternos */
proc freq data=sire4 noprint; tables name
/out=count norow nocol nocum nopercent;
run;
proc sort data=sire4;by name;run;
proc sort data=count;by name;run;
data sire4; retain count;
  merge count sire4; by name;
  Avocount=count;
  drop percent count;
run;
proc sort data=sire4 nodupkey; /* elimina duplicatas */
  by name;
run;
proc sort data=sire3 nodupkey; /* elimina duplicatas */
  by name;
run;
/*Junta pais e avos paternos e elimina avos com somente 1 filho e que
não
 aparecem como pai de outros produtos */
data sire5;
  merge sire3(in=temp) sire4(in=temp1 rename=(animal=hbbavo));
  by name;
  if temp=0 & avocount=1 then delete;
  if animal='' then animal=hbbavo;
  if hbb_pai^='';
rm=0;
  touro=hbb_pai;
  keep animal touro rm;
run;
data ped3; /* acrescenta touros com pai conhecido como produtos no
```

```
arquivo de pedigree */
  set ped3 sire5;
  keep animal touro vaca rm;
run;
proc sort data=ped3 nodupkey; by animal;
run;
data touro;
  set ped3(where=(rm=0)); /* le os touros*/
  animal=touro;
  keep animal;
run;
proc sort nodupkey; by animal;       /* elima duplicatas */
run;
* FIM DO OPCIONAL;
data vaca;
  set ped3;                        /* le as vacas */
  animal=vaca;
  keep animal;
run;
proc sort nodupkey; by animal;       /* elima duplicatas */
run;
data animal;
  set ped3 touro vaca;
  if animal^='';
  keep animal;
run;
proc sort nodupkey; by animal;       /* elima duplicatas */
run;
proc sort data=ped3 nodupkey; by animal;       /* elima duplicatas */
run;
/* renumera animais (inclusive touros e vacas) de 1 a n de animais */
data ped3;
  merge animal(in=temp) ped3;
  by animal;
  if temp=1;
  ianimal+1;
keep animal touro vaca rm ianimal;
run;
/* Junta as identificação original e recodificada de touros*/
data touro;
```

```
set ped3;
  touro=animal;
  keep touro ianimal;
run;
proc sort data=ped3; by touro; run;
proc sort data=touro nodupkey; by touro; run;
data ped3; /* renumera touros */
  merge ped3(in=temp) touro(in=temp1 rename=(ianimal = id));
  by touro;
  if temp=1;
  if temp1 =1 then itouro=id;
  drop id;
run;
/* Junta as identificação original e recodificada de vacas*/
data vaca;
  set ped3;
  vaca=animal;
  keep vaca ianimal;
run;
proc sort data=ped3; by vaca; run;
proc sort data=vaca nodupkey; by vaca; run;
data ped3; /* renumera vacas */
  merge ped3(in=temp) vaca(in=temp1 rename=(ianimal = id));
  by vaca;
  if temp=1;
  ivaca=id;
  if itouro =. then itouro =0;
  if ivaca =. then ivaca =0;
  drop id;
run;
/* Grava o arquivo de pedigree */
proc sort data=ped3; by animal; run;
data _null_;
  set ped3;
  file "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\Angusgpd.ped";
  if (ivaca=0 & itouro=0) then code=3;
    else if (rm=1 & ivaca>0) then code=2;
    else if (itouro>0 & ivaca>0) then code=1;
          else code=2;
```

```
 put ianimal itouro ivaca code animal touro vaca;
run;
/* 3. Arquivo de dados para modelo animal */
/* Renumera grupos de contemporâneos de 1 a n*/
proc sort data=gpd3;  by gc; run;
data rec;
  set gpd3(drop=igc);
  by gc;
  if first.gc then igc+1;
run;
* Junta a numeração recodificada dos animais no arquivo de dados;
proc sort data=rec;
  by animal;
run;
proc sort data=ped3;
  by animal;
run;
data rec;
  merge rec(in=temp) ped3;
  by animal;
  if temp=1;
run;
proc sort data=rec;
  by ianimal;
run;
* Grava arquivo de dados;
data _null_;
  set rec;
  file "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\Angusgpd.dat";
* listar aqui o animal (vaca), efeitos fixos e aleatórios e variáveis
resposta a serem incluidos no modelo. O 1 é usado aqui para colocar
uma média geral no modelo, mas é opcional;
  put ianimal ' 1 ' igc
                    CSP IDV IDV2 IDS IDS2
             gpd;
run;
* 4. Estatisticas descritivas para conferir o arquivo de dados gravado.
     Lê arquivo gravado;
data dat;
```

```
infile "C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\Angusgpd.dat";
   input animal mu igc
                   CSP IDV IDV2 IDS IDS2
            gpd;
run;
proc means data=dat;
  var animal mu igc
                   CSP IDV IDV2 IDS IDS2
            gpd;
run;
proc freq data=dat;
  tables igc csp idv ids/nocol norow nocum nopercent;
run;
* 5. Roda modelo fixo para testar relevância dos efeitos no modelo;
proc mixed data=dat;
  class IGC;
  model gpd = IDV IDV2 IDS IDS2
           /solution;
  random igc;
run;
```

Anexo 6 – Pré_INTERGEN3.sas:

```
* Pre_INTERGEN3.sas Rotina de preparação do arquivo para rodar o
modelo de normas de reação
* no programa INTERGEN, para estudos de interação genótipo-ambiente
* IMPORTANTE: requer que a rotina Pre_INTERGEN2.sas tenha sido
usada para gerar o arquivo de
* dados e que o programa INTERGEN tenha sido rodado com o modelo
animal nesses dados
* Por Fernando F. Cardoso 2006.
* Dados Angus do PROMEBO pós-desmama arquivo extensão .INS ;
options nodate;
title1 'Pre_INTERGEN3 - Promebo Angus';
%let dir = C:\Documents and Settings\fernando\My
Documents\INTERGEN\Angus\PD\;
%let file = Angusgpd.dat;
* 1. Lê arquivo de dados gravado pela rotina Pre_INTERGEN2.sas;
data dat;
```

```
infile "&dir&file";
  input animal mu igc
                CSP IDV IDV2 IDS IDS2
          gpd;
run;
* 2. Lê arquivo "solutions" com as soluções do programa INTERGEN
para os efeitos
  fixos e aleatórios do modelo animal
  IMPORTANTE: É preciso especificar qual entre os efeitos contém as
soluções para os efeitos
  ambientais - neste caso são os grupos de contemporâneos - efeito 3 -
na terceira linha da
  secção EFFECTS do arquivo de parâmetros para rodar modelo animal
no INTERGEN;
%let file = Ma_gca\solutions;
%let effect = 3; * número do efeito no arquivo solutions;
%let x = gc; * indica que o efeito ambiental (x) será chamado de gc;
data sol;
  infile "&dir&file";
  input trait effect level mean sd last;
  if trait^=.;
  if effect=&effect;
  i&x=level;
  sol_&x=mean;
  keep i&x sol_&x;
run;
/*
* OPCIONAL QUANDO HOUVER MAIS DE UMA CADEIA;
%let file = Ma_gcf\solutions;
%let x = gcf; * indica que o efeito ambiental (x) será chamado de gc;
data sol2;
  infile "&dir&file";
  input trait effect level mean sd last;
  if trait^=.;
  if effect=&effect;
  i&x=level;
  sol_&x=mean;
  keep i&x sol_&x;
run;
data sol;
```

```
  merge sol sol2;
run;
proc univariate plot;
var sol_gcf sol_gc;
run;
proc corr;
  var sol_gcf sol_gc;
run;
proc gplot;
  plot sol_gcf*sol_gc;
run;
proc reg;
  model sol_gcf=sol_gc;
  plot sol_gcf*sol_gc;
run;
*/
proc univariate plot;
  var sol_&x;
run;
proc sort data=dat;
  by i&x;
run;
* 3. Junta dados e soluções para o efeito ambiental e cria polinomios de
Legendre de ordem
* o, especificada abaixo (o=1 -> linear, intercepto e inclinação, o=2 -
> quadratico, etc)
Note que se a regressão aleatória é linear (o=1), como tipicamente em
modelos de normas de
reação, não há necessidade de utilizar polinomios de legendre;
%let o = 1; * Ordem dos polinomios de legendre;
* Obtem minimos e máximos da variável sol_&x;
proc means data=sol min max noprint;
  var sol_&x;
  output out=m&x min=min&x max=max&x;
run;
proc print; run;
data iga;
  merge dat(in=temp) sol;
  by i&x;
  if temp=1;
```

```
  if _n_ = 1 then set m&x; * insere min e max of &x em todos os
registros;
  * Polinomios de Legendre;
  std&x=-1+2*((sol_&x-min&x)/(max&x-min&x)); *Padroniza valores de
&x entre -1 e 1;
  array p{0:&o}p0-p&o; * função P - ver L. Schaeffer Notes ANSC637
Set 14 Random Regr. Models;
  array f{0:&o}f0-f&o; * polinomios de Legendre de ordem &o;
  p0=1; p1=std&x;
do n=0 to &o;
   if n>0 and n<&o then
          p{n+1}=(1/(n+1))*((2*n+1)*std&x*p{n}-n*p{n-1});
        f{n}=((n+.5)**.5)*p{n};
  end;
  output;
run;
* Checa se polinomios gerados correlações devem ser =1;
proc corr data=iga;
  var f1 p1 sol_&x;
run;
* 4. Gera arquivo de dados para rodar modelo de normas de reação;
%let file = Angusgpd_iga.dat;
proc sort data=iga;
  by animal;
run;
data _null_;
  set iga;
  file "&dir&file";
  put animal ' 1 ' igc sol_gc f0-f&o
                    CSP IDV IDV2 IDS IDS2
            gpd;
run;
```

Anexo 7– Pós_INTERGEN1.sas:

```
* Pos_intergen1(CompVar).sas - Rotina para análise do pós-gibbs do
arquivo de saida dos componentes de variancia do programa Intergen
(varcompsam), para verificar convergencia e para inferencia a partir da
cadeia de Gibbs.
  Por Fernando F. Cardoso 2006.
```

```
Macro var(direct=diretorio onde se encontram as analises, analysis=
nome do modelo ou analise, burnin= numero de ciclos de aquecimento,
skip= intervalo de thinning, nvc= numero de componentes de
  variancia na analise) Macro var1(par=parametro para analise conforme
nomes de variaveis na macro var, nlag= atraso maximo para calcular a
autocorrelacao no PROC ARIMA que garante que esta chegue a zero);

options nodate ps=66 ls=80;
data varf;
run;
data dist;
run;

symbol1 value = none color = red interpol = join line = 1 h=.2;
symbol2 value = none color = blue interpol = join line = 1 h=.2;
symbol3 value = none color = green interpol = join line = 1 h=.2;
%macro var(direct,analysis,burnin,skip,nvc);
title1 "Componentes de Variancia - &analysis";
DATA varsf;
  * Le arquivo de saida do Intergen;
  %let file = &direct\varcompsam;
  INFILE "&file";
  INPUT cycle;
  model="&analysis";
  %do i=1 %to &nvc;
    INPUT vc&i;
    svc&i+vc&i; * calcula a soma cumulativa das amostras de cada
compvar;
  %end;
  if mod(cycle,&skip)=0;
RUN;
title2 "Correlacao das amostras dos diferentes componentes de
variancia";
* Calcula correlacao das amostras por ciclo dos diferentes compvar;
PROC CORR data=varsf;
   where cycle > &burnin;
   var vc1-vc&nvc;
RUN;
```

```
* Calcula a media pos burnin de cada comp var;
PROC MEANS data=varsf mean noprint;
   where cycle > &burnin;
   var vc1-vc&nvc;
   output out=mvar mean=mean_vc1-mean_vc&nvc
RUN;

* Cria variaveis necessarias para calcular autocorrelacao das amostras
de cada compvar
  com atrasos(lag) de 1, 10 e 100 ciclos;
DATA varsf;
  merge varsf
    varsf(firstobs=2 rename=(vc1-vc&nvc=o2vc1-o2vc&nvc)
drop=cycle svc1-svc&nvc)
    varsf(firstobs=2 rename=(svc1-svc&nvc = s2vc1-s2vc&nvc)
drop=cycle vc1-vc&nvc)
    varsf(firstobs=11 rename=(vc1-vc&nvc=o11vc1-o11vc&nvc)
drop=cycle svc1-svc&nvc)
    varsf(firstobs=11 rename=(svc1-svc&nvc = s11vc1-s11vc&nvc)
drop=cycle vc1-vc&nvc)
    varsf(firstobs=101 rename=(vc1-vc&nvc=o101vc1-o101vc&nvc)
drop=cycle svc1-svc&nvc)
    varsf(firstobs=101 rename=(svc1-svc&nvc = s101vc1-
s101vc&nvc) drop=cycle vc1-vc&nvc)
    varsf(firstobs=102 rename=(vc1-vc&nvc=o102vc1-o102vc&nvc)
drop=cycle svc1-svc&nvc)
    varsf(firstobs=102 rename=(svc1-svc&nvc = s102vc1-
s102vc&nvc) drop=cycle vc1-vc&nvc)
;
  if _n_=1 then set mvar;
* calcula media corrente lag=100 - nao muito interessante!;
  %do i=1 %to &nvc;
     av100vc&i = (s102vc&i + s101vc&i - s2vc&i - svc&i)/200;
  %end;
RUN;

DATA var;
  set varsf;
  if cycle > &burnin;
  * Componentes de variancia adicionais das normas de reacao
```

```
 IMPORTANTE: ISTO NÃO É GERAL E PRECISA SER CONFERIDO EM
CADA ANÁLISE;
  if model^='ma_gca' then /* esta linha deve indicar quando for modelo
animal padrao */
    do; * nao faz se for modelo animal;
          * Correlacao entre nivel/intercepto e inclinação da norma de
reação;
            corab=vc3/(sqrt(vc2)*sqrt(vc4));  * Neste caso vc2 é
var(nivel), vc4 é var(inclinacao)

  e vc3 é covariancia entre nivel e inclinacao;
    * Variancia genetica em determinado ambiente (x)
            varg|x=var(nivel)+x**2*var(incl)+2*x*covar(nivel,incl);
            x=50; *valor de x;
      varg_b=vc2+x**2*vc4-2*x*vc3; * -x (ambiente baixo);
      varg_a=vc2+x**2*vc4+2*x*vc3; * x (ambiente alto);
      h2 =vc2/(vc2+vc5); * herdabilidade quando x=0 (ambiente medio)
vc5=var(erro);
      * Ajusta var(erro) em modelos heteroscedasticos (h) no gradiente
ambiental;
      if model='mhnrh' | model='mhnrh1' then
        do;
        * Var(erro) e herdabilidade em determinado ambiente (x);
                vare_b=vc5*(vc6**(-x));
                vare_a=vc5*(vc6**x);
          h2_b  =varg_b/(varg_b+vare_b);
          h2_a  =varg_a/(varg_a+vare_a);
              end;
             * No caso de erros homoscedasticos;
      else
        do;
                  * Herdabilidade em determinado ambiente (x);
          h2_b  =varg_b/(varg_b+vc5);
          h2_a  =varg_a/(varg_a+vc5);
              end;
           end;
  else
           h2  =vc1/(vc1+vc3); * herdabilidade para modelo animal
padrao;
RUN;
```

```
%do i=1 %to &nvc;
* Cria graficos de traco dos compvar e calcular autocorrelacoes de
atraso 1, 10 e 100;
title2 "Graficos das Amostras e Media";
  proc gplot data=varsf;
    plot vc&i*cycle=1 mean_vc&i*cycle=2
/*av100vc&i*cycle=3*//overlay legend;
  run;
PROC CORR data=var;
title2 "Autocorrelacao(1) das Amostras";
   var vc&i o2vc&i;
RUN;
PROC CORR data=var;
title2 "Autocorrelacao(10) das Amostras";
   var vc&i o11vc&i;
RUN;
PROC CORR data=var;
title2 "Autocorrelacao(100) das Amostras";
   var vc&i o101vc&i;
RUN;

* Chama macro var1 para cada varcomp;
%var1 (vc&i.  ,1000);

%end;

%var1 (h2    ,1000);

* Componentes de variancia adicionais das normas de reacao;
%if &analysis ^= ma_gca %then %do; * nao faz se for modelo animal;
    %var1 (corab,1000);
    %var1 (varg_b,1000);
    %var1 (varg_a,1000);
          %if &analysis=mhnrh | &analysis=mhnrh1 %then %do;
            * Componentes modelos heteroscedasticos;
              %var1 (vare_b,1000);
              %var1 (vare_a,1000);
        %var1 (h2_b,1000);
        %var1 (h2_a,1000);
            %end;
```

```
%else %do;
            * Componentes modelos homoscedasticos;
        %var1 (h2_b,1000);
        %var1 (h2_a,1000);
            %end;
  %end;

%mend var;

* Estimativa da densidade marginal dos parametros, incluindo media, moda, media e
  percetis de 2.5 e 97.5, variancia de monte carlo e numero efetivo de amostras;
%macro var1(par,nlag);
proc means data=var noprint;
  var &par;
  output out=mean mean=mean std=std;
run;
* Densidade marginal dos parametros, incluindo media, moda, media e
  percetis de 2.5 e 97.5;
proc kde data=var out=dist&par percentiles=2.5 50 97.5
method=snr; * bwm=2;
  var &par;
  ods output percentiles=perc;
run;

proc means data=dist&par noprint;
   var density;
   output out=mode maxid(density(&par))= mode;
run;

data dist&par;
  set dist&par(rename=(&par=variance density=&par)) ;
  model="&analysis";
run;

data dist;
  set dist dist&par;
run;
```

```
proc arima data=var;
  identify var=&par nlag=&nlag outcov=corr noprint;
run;
*proc print data=corr;
*run;

* Variancia de Monte Carlo e numero efetivo de amostras;
proc iml;
 use corr;
 read all var{'LAG'}into lag;
 read all var{'CORR'}into corr;
 read all var{'COV'}into cov;
 read all var{'n'}into n;
 nsample = max(n);
 free n;
 nlag = nrow(lag);
 nlag2 = nlag/2;
 Gamma = j(nlag2,1,0);
 cutoff = 0;
 do t = 1 to nlag2;
   Gamma[t] = cov[2*(t-1)+1] + cov[2*(t-1)+2];
   if t > 1 then
     if Gamma[t] > Gamma[t-1] then Gamma[t] = Gamma[t-1];
   if Gamma[t] < 0 then cutoff = 1;
   if cutoff = 1 then Gamma[t] = 0;
 end;
 varm1 = (-cov[1] + 2*sum(Gamma)) / nsample; * variancia de Monte
Carlo;
 ess1 = cov[1]/varm1;  /* Numero efetivo de amostras - effective
sample size */
 stdm1 = sqrt(varm1);          /* Desvio padrao de Monte Carlo  */
* print cutoff stdm1 ess1;

/* Bacthing method */
 use var;
 read all var{"&par.    "}into var;
 m = nrow(var); * number of samples;
 n=0;                     * batch size;
 r=1;                     * lag-1 autocorr;
 do while (n < m/10 & r > .05);
```

```
   n=n+10;
   k=floor(m/n);
   bmean=j(k,1,0);      *batch means;
   do i=1 to k;
           bmean[i]=var[(i-1)*n+1:(i*n)][:];
   end;
   mu=bmean[:];
   v=(bmean[##]-bmean[+]**2/k);
   r=((bmean[1:k-1]-mu)`*(bmean[2:k]-mu))/v;
 end;
 if r <= .05 then
   do;
     print n k r;
           varm2 = v/(k*(k-1));
           ess2 = ((var[##]-var[+]**2/m)/m)/varm2;  /* effective sample
size */
           stdm2 = sqrt(varm2);          /* Monte Carlo standard deviation
*/
*          print stdm2 ess2;
   end;
 else
   do;
           ess2 = .;  /* effective sample size */
           stdm2 = .;     /* Monte Carlo standard deviation  */
           print 'batching failed to reach r <=.05';
   end;

 meth="&analysis.  ";
 par="&par.    ";
 create ess1 var{meth par ess1 ess2 stdm1 stdm2 cutoff n r};
 append;
run;
quit;

data var1;
  merge  mean(keep=mean std)
                              mode(keep=mode)
                              perc(where=(percent=50) rename=(&par  =
p500))
             perc(where=(percent=2.5) rename=(&par  =  p025))
```

```
                              perc(where=(percent=97.5) rename=(&par =
p975))
                              ess1;
  drop percent;
  run;

data varf;
  set varf var1;
run;
%mend var1;
%var(C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\mhnrh_restart,mh
nrh,10000,10,6);
%var(C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\mhnrh1,mhnrh1,
10000,10,6);
%var(C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\ma_gca,ma_gca,1
0000,10,3);

%let direct = C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\;
proc print data=varf; run;
PROC EXPORT DATA= WORK.varf
            OUTFILE= "&direct.lga_Angus_CompVar"
            DBMS=EXCEL2000 REPLACE;

RUN;
PROC EXPORT DATA= WORK.dist
            OUTFILE= "&direct.lga_Angus_PDist"
            DBMS=EXCEL2000 REPLACE;
RUN;
```

Anexo 8– Pós_INTERGEN2.sas:

```
* Pos_Intergen2.sas - Rotina de macro para implementação de criterios
de escolha de modelos, a partir das saídas do Programa Intergen
(loglike_rnd e loglike_obs).
  Por Fernando F. Cardoso 2007.
```

```
Macro model(direct=diretorio onde se encontram as analises, analysis=
nome do modelo ou analise, burnin= numero de ciclos de aquecimento)
;
options nodate ps=66 ls=80;
/* Model Choice Program CPO/PBF, BF and DIC */
data model;
run;
%macro model(direct,analysis,burnin);
title1 "Model Choice Criteria - &direct ";
DATA lobs&analysis;      /* input mean likelihood values for each obs */
  %let file = &direct.loglike_obs;
  filename in "&file";
  INFILE in;
  analysis="&analysis";
  INPUT obs cpo&analysis devi&analysis devmi&analysis ei&analysis;
  if _n_<=2 then delete;
  devi&analysis=-2*devi&analysis;
  devmi&analysis=-2*devmi&analysis;
  pdi&analysis=devi&analysis-devmi&analysis;
  dici&analysis=devi&analysis+pdi&analysis;
RUN;
DATA lrnd&analysis;      /* input likelihood values for each cycle */
  %let file = &direct.loglike_rnd;
  filename in "&file";
  INFILE in;
  INPUT round bdev&analysis;
  bdev&analysis=-2*bdev&analysis;
  if round > &burnin;
RUN;
/* DIC - Deviance Information Criteria */
PROC IML;
use lobs&analysis;
read all var {cpo&analysis}into cpo;
read all var {devi&analysis}into devi;
read all var {devmi&analysis}into devmi;
close lobs&analysis;
use lrnd&analysis;
read all var {bdev&analysis}into bdev;
close lrnd&analysis;
Dev=devi[+]; *Dev=bdev[:]; *another way to calculate dev;
```

```
pD=Dev-devmi[+];
DIC=Dev+pD;
CPOdev = -2*(sum(log(cpo)));
/* BF - Bayes Factor */
k=max(bdev);
BFdev=0;
do i=1 to NROW(bdev);
  BFdev=BFdev+exp(bdev[i]-k);
end;
BFdev=log(BFdev/NROW(bdev))+k;

repl="&analysis";
create model1 var{repl Dev pD DIC CPOdev BFdev};
append;
quit;
/* create data sets */
data model;
  set model model1;
run;

proc kde data=lrnd&analysis out=dist_bdev&analysis  method=snr
percentiles=2.5 50 97.5;
  var bdev&analysis;
run;

%mend model;

%model(C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\mhnrh1\,mhnrh1,
10000);
%model(C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\mhnrh_restart\,m
hnrh,10000);
%model(C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\ma_gca\,ma_gca,
10000);
title2 ' Model Choice Criteria - DIC, PBF & BF';
proc print data=model; run;
```

```
%let direct = C:\Documents and Settings\TOSHIBA\My
Documents\Embrapa\Projetos\Intergen\Angus\GPD345\;
PROC EXPORT DATA= WORK.model
          OUTFILE= "&direct.Iga_Angus_Model"
          DBMS=EXCEL2000 REPLACE;
RUN;
```